ARMY RESEARCH LABORATORY

Aviator Behavior and Performance as Affected by Aircrew Life Support and Protective Equipment

John D. Waugh
Linda T. Fatkin
Debra J. Patton
Linda L. Mullins
Pamela A. Burton
Daniel J. Barker
David A. Mitchell

ARL-MR-440 MARCH 1999

Army Research Laboratory
Aberdeen Proving Ground, MD 21005-5425

ARL-MR-440 March 1999

Aviator Behavior and Performance as Affected by Aircrew Life Support and Protective Equipment

John D. Waugh
Linda T. Fatkin
Debra J. Patton
Linda L. Mullins
Daniel J. Barker
Human Research & Engineering Directorate, ARL

David A. Mitchell
Arizona Army National Guard

Abstract

A methodology for quantifying Army rotary wing aviator performance as influenced by aircrew life support, survival, and nuclear-biological-chemical clothing and equipment ensembles was examined in a set of experimental trials conducted in an AH-64 (Apache) combat mission simulator. The methodology was based on an aircrew evaluation procedure originally developed for use in the crew coordination training of all Army aviators. It uses a set of 13 basic qualities, each with behavioral anchors and a 7-point rating scale, and it is administered by specifically trained senior aviator evaluators. Ten crews, two aviators in each, while fully encumbered, performed three combat missions for record, representative of typical operational tasks, with one "variation" trial conducted without the over-water components of the ensemble. Measures of effectiveness and flight data, as well as stress assessment and equipment "complaints" citations, were recorded. The results indicated that the behavior-anchored scores were not sensitive enough to statistically discriminate among the independent variables of repeated measures and the variation trials even though graphically, differences were readily apparent. Attempts to apply transformations to the data, based on the aviator subjects' relative flying experience and their apparent accommodation to the trials were also statistically unsuccessful. The additional measures collected did not yield statistically significant discriminations nor did they correlate well with the evaluation scores. A number of options for improving the technique are offered.

ACKNOWLEDGMENTS

The authors offer thanks to all the people who helped in this project. From the Human Research and Engineering Directorate (HRED) of the U.S. Army Research Laboratory, Mr. Gordon Herald, for his expertise in computers, modern and not so modern, and to Mr. Thomas Davis and Mr. Judah Katznelson for assistance in data reduction. From the U.S. Army Aviation Warfighting Center, Ft. Rucker, Alabama, Ms. Laura Thompson and CW4 Larry Thompson, for helping us understand the AH-64 combat mission simulator (CMS). From the HRED Field Element, Ft. Rucker, Mr. Richard Armstrong and Mr. David Durbin, for their technical reviews and all-round assistance in all phases of this project. From the Army Research Institute (ARI), also in Ft. Rucker, Dr. Robert Wright, for sharing insights and data from ARI's allied investigations. To CW5 Kenneth Donahue, Training and Doctrine Command Systems Manager Longbow, Ft. Rucker, for his technical review, and to MAJ Steve Walters and CW4 Steve Paris, U.S. Army Aviation Technical Test Center, Ft. Rucker, for working closely with us in the initial survey of aircrew training manual (ATM) tasks. To CW5 Thomas Connell and CW5 Howard Swaim, Flight Standardization, Ft. Hood, Texas, for their advice, counsel, and assistance in our survey of ATM tasks.

We also offer an appreciative and hearty "well done" to that highly motivated group of associate investigators who were major contributors to this investigation by working long, long hours during the trials while maintaining their sense of humor throughout: Ms. Jean Breitenbach and Mr. Ronald Whittaker as well as co-authors Mr. Daniel Barker, Ms. Debra Patton, and Ms. Linda Mullins, HRED; and to those highly skilled instructor-operators and evaluators manning the simulator, CW3 C. David Pope, Jr., CW4 Michael Hillwig, CW5 Gary Kilker, Mr. Charles (Chip) Decker, WO2 Joseph O'Neill, WO2 Michael Wanamaker. Of course, we are extremely grateful to the aviators who participated as subjects, whom we cannot thank by name in this report. Nevertheless, they are highly commended for their dedication to the investigation as well as their willingness to tolerate strange hours and large doses of hard work and discomfort.

Special gratitude is given to the people at the Western Area Aviation Training Site (WAATS), Marana, Arizona, who most graciously offered their simulation facilities for the experiment and who assisted, along with the Arizona National Guard, in the logistics of its execution: COL George Gluski, CDR, WAATS, LTC James Braman, Chief of the AH64 CMS facility; his second in command and co-author MAJ David Mitchell; Platoon Leader and master of the schedule CW4 Don Curry; WO2 William Mercer who made the right things happen, day after day after day, and the whole WAATS staff who cheerfully saw to our many needs and logistics for the entire 6 weeks worth of trials. Special thanks to MAJ Kenneth Nettles, CDR Arizona Army National Guard (ARNG) Army Aviation Support Facility, Silver Bell Army Heliport (next door to WAATS) and his staff for supplying participants, equipment and all manner of support, often on a short fuze basis.

Also, most worthy of mention are the people we only knew via the telephone but without whom we would not have been able to conduct the project: From Headquarters, National Guard Bureau, MAJ Brian West who secured the National Guard's approval to employ the WAATS facilities and broadcast our need for participants, and Mr. Howard Manwieler who obtained the

pay and allowance dollars to support those participants. To Ms. Colleen Heck, ARNG Budget Officer (Phoenix) who performed all the financial wizardry between the National Guard Bureau, HRED, and the participants. To the various state National Guard points of contact who did recruiting as well as arranging, scheduling, and the paper work for the participants: CW3 Russell Thacker, Utah ARNG; CPT Dale Hall, South Carolina ARNG; WO2 David Kelcevick, North Carolina ARNG.

John D. Waugh
Principal Investigator

CONTENTS

FIGURES

TABLES

EXECUTIVE SUMMARY

This report describes an experiment conducted by the Human Research and Engineering Directorate of the U.S. Army Research Laboratory (ARL) in support of the Project Manager, Air Crew Integrated Systems. The purpose of the experiment was to conduct research into the development of a technique for use in the assessment of future rotary wing aviation aircrew life support (ALSE) ensembles and components, including survival and nuclear-biological-chemical (NBC) clothing and equipment.

The experiment consisted of a set of trials with ten male AH-64 Apache crews flying tactical, combat missions in the AH-64 combat mission simulator located at the Western Area Aviation Training Site, Marana, Arizona. The crews were outfitted in the Army's current inventory of NBC, and ALSE, in missions described as hostile threat, over-water survival, and NBC environments.

Each crew flew four trials or missions. The first trial was a familiarization trial done primarily to bring the subjects into the condition of having to don and perform with the entire collection of equipment never before experienced. Three trials for record were then conducted, two for a repeated measures application and the third trial being a "variation" trial where the crews did not don the over-water components of the ensemble. The three trials for record were counterbalanced in their order of presentation to the crews. The mission scenario for the three trials for record was the same for each crew and each trial. The familiarization trials were all done with a separate scenario and terrain.

The primary measure of individual crew member's performance was an adaptation of the successful crew coordination evaluation procedure developed for Army Crew Coordination Training in 1995-1996 by the Army Research Institute for the Behavioral Sciences. The technique is based on 13 sets of "behavioral anchors," for each of 13 basic qualities against which crews are evaluated at the end of their crew coordination training. Aviators are rated along a 7-point scale from "very poor" through "acceptable" to "superior". In this experiment, individual performance was scored by certified evaluators, not against the individual himself, but against the effects of the encumbering ensemble on the individual, which could possibly degrade his performance. To that end, the behavioral anchors were edited to make the ensemble the object of the evaluator's judgment. A number of other measures were recorded concurrently, including measures of effectiveness and flight data from the simulator, pre- and post-trial questionnaires and interviews, and an assessment of the stress levels present in the subjects by a battery of

questionnaires and the assay of salivary amylase, all intended to corroborate the mainline evaluation findings.

The evaluation scores that were gathered did not have sufficient sensitivity to the experimental independent variables for two fundamental reasons: first, too large a variation was present from crew to crew in the scores, well exceeding the span of a single point of the 7-point rating scale. Second, the range of awarded scores was too narrow in terms of its overall distribution. The end points of the scale never appeared in the scores posted in the trials; they all clustered about the mid-point, between 4 (acceptable) and 5 (good) on the scale. Some of the other measures collected showed statistical significance with respect to the experiment's conditions, usually the second of the two fully encumbered trials and the variation trial. Some systemic problems were encountered in gathering those data, but then, the data were not intended to be subjected to the degree of statistical scrutiny reserved for the primary measure. The aviators made a great number of citations where the ensemble components caused problems or subjectively caused their performance to be degraded or contributed significantly to their discomfort. Principally, the protective mask, followed by the body armor and protective overgloves were cited, as well as combinations of component such as helmet and mask or mask and armor. In the copilot-gunner's position, the body armor in front plus the life raft on the back physically prevented full aft cyclic displacement from either seat. The collection of stress assessment measures showed that the aviators experienced significantly more stress in performance of their missions versus pre-trial and recovery after-trials measures, but the level of stress was not as great as that encountered by participants in other investigated situations. A moderate amount of psychological stress was concluded to be present. When the subjects were encumbered in ALSE in a low physical work environment, physiological stress created by the ensembles was minimal with respect to that in other investigations.

While not statistically viable in this experiment, graphically, it is very apparent that the subjects improved their performance from the first through the third trials for record in sequence. Rather than calling this improvement "learning" in the classical sense, "accommodation" is the term of choice because the subjects already have learned to fly, fight, and work in ALSE. A most puzzling observation of the evaluation scores revealed that the subjects performed more poorly during their variation trials, when they wore less equipment and no over-water gear, than during the second trial for record (recall that these trials were counterbalanced for order of presentation). It was fully expected that they would perform best in this trial, but it was not the case. It was also thought that the more experienced pilots (two had at least 5000 total hours logged) would perform at a higher level than their less experienced brethren. This was not the case either, and the notion of

perhaps weighting the scores in this fashion to potentially increase their sensitivity had to be abandoned. Adjusting the scores to remove the influence of the suggested accommodation effect appeared to increase their sensitivity but not quite enough to bring them into statistical significance.

This is not to say that the concept of employing behaviorally anchored performance evaluations as a tool will not work for assessing the merit of prototype or proposed future ALSE ensembles or components. Rather, that the technique should be modified in order to eliminate those properties that caused its insensitivity. Recommended solutions include reduction of the 7-point scale to a 5- or even a 3-point scale, leaving less room for variation and encouraging the certified evaluators to be rather ruthless in their evaluations and to be willing to use the end points of the scale more so than was observed in this set of trials. The adjusting of scores to cancel the conjectured performance increases from trial to trial should be continued at least until we have more knowledge of the aviator's ability to accommodate to the fully encumbered environment. Finally, future trials should be conducted, probably with proposed ensembles or components on a comparative basis, evaluating missions against the current ALSE that would be replaced. That way, a current equipment baseline can be further defined, and meaningful assessments of advanced technology equipment can be conducted. Include the technique(s) that may be available to assess the relative levels of the aviator's situation awareness as another potentially useful factor.

AVIATOR BEHAVIOR AND PERFORMANCE AS AFFECTED BY AIRCREW LIFE SUPPORT AND PROTECTIVE EQUIPMENT

INTRODUCTION

Military rotary wing aviators are required, depending on their mission, to don a large variety of life support and protective equipment and components that are intended to increase their probability of survival in tactical situations. Among this equipment are items for survival in the event of over-water "ditching," protection against nuclear-biological-chemical (NBC) agents, enemy fire, crashing, fire, eye-damaging lasers, and extreme temperatures, hot and cold. Add to this equipment an emergency locator transmitter, a side arm, survival knife, flashlight, and so forth. The average weight of the standard flight uniform, clothing, helmet, and boots is 17.7 lb, and the average weight of a mission-oriented protective posture (MOPP) IV fully encumbered ensemble is 57.1 lb (Reardon et al., 1996).

The investigation cited above concluded that discomfort experienced by aviator participants, in terms of increased core temperatures and dehydration because of increased sweat rates and equipment pressure points, affected mood and mission endurance. An investigation in a simulator emulating the RAH-66 Comanche (Sharkey & Schwirzke, 1995) made similar conclusions and highlighted a significant decrement in the time it took to perform data entry tasks in MOPP IV. A report of the man-machine integration design and analysis system (MIDAS) modeling investigation of the AH-64D Longbow Apache cockpit (Shively et al., 1995), which presented a scenario of flying and fighting tasks for the Longbow Apache mission in concert with employment of the Jack® human figure model (no human subjects), stated that the results "... clearly show some of the performance decrements associated with today's protective ensemble." Finally, a U.S. Army Research Institute (ARI) investigation (Wright, Hartson, & Couch, 1996) in an AH-64 configured simulator cited the same discomfort, heat stress, "hot spot" pressure points and endurance difficulties, performance decrements in tasks as well as equipment-cockpit interferences.

Universal acknowledgment of the problems associated with aviators and ALSE has led to the establishment of the Project Manager for Air Crew Integrated Systems (PM ACIS) and the Air Warrior Program described as "... the rotary wing aviation focus for providing a mission-tailorable system that standardizes and integrates Aviation Life Support Equipment for Army, Navy, and Marine Corps aircrew personnel during flight and ground operations" (Metzler, 1995). The PM ACIS has retained the Human Research and Engineering Directorate (HRED) of the U.S. Army Research Laboratory (ARL) to develop the capability to conduct early engineering analyses

of emerging technologies and designs aircrew life support equipment (ALSE) and protective components[1] and to estimate any potential payoff in mission performance improvement. This capability is expected to allow both the user proponency and materiel developers to better standardize and integrate all this equipment as a soldier-based system for the rotary wing aviation mission. The ARL focus was on the AH-64 Apache as the basis of investigations because it represents the platform with the most confined cockpit space and highest workload; the knowledge gained could potentially be extrapolated to the AH-64D Apache Longbow, successor to the Apache.

The investigation described herein is a fundamental research effort for the Air Warrior Metrics program performed on behalf of the PM ACIS. In addition to becoming a behavioral and performance baseline for ALSE affects, the results of the Air Warrior Metrics investigations are also expected to contribute to improved performance research integration tool (IMPRINT) analyses of current and future helicopter systems.

OBJECTIVE

The object of this investigation will be to develop performance-based measurement scales which may then be used to estimate the merits (or lack thereof) of future ALSE components and their use on rotary wing aircrews. The assessment of aircrew ensemble technologies should be provided in a standardized framework that will allow the accumulation of a database across airframe systems and ALSE systems and components, which is keyed to soldier (human) performance. To be useful, these measurement scales should be (a) developed around the intended mission(s) for each current or proposed helicopter type, as defined in its aircrew training manual (required aviator and aircrew tasks), (b) sensitive to ALSE technologies and designs, (c) easily observable and measurable, and (d) have demonstrated validity and reliability.

The helicopter agreed upon by PM ACIS and ARL to be the initial platform to be studied is the attack helicopter AH-64 Apache. The required aviator and aircrew tasks in the AH-64 aircrew training manual (ATM) assigned to this aircraft (Headquarters, Department of the Army, 1992) represent both flying and fighting tasks to be performed by a two-person crew, both rated aviators, positioned in tandem with the copilot-gunner (CPG) in the forward seat.

---

[1]Use of the acronym ALSE will imply all life support, survival, and MOPP equipment and components for the remainder of this report.

The total number of tasks listed in the ATM is 88, but not all these tasks were necessarily exercised in the investigation. Preflight, postflight, and certain redundant tasks have been eliminated. Further culling of tasks has been accomplished by the surveying of subject matter experts (SMEs) before experimentation (see Method and Procedure). It was also agreed that the investigation and subsequent measurement scales consider as the definitive baseline, the condition of aviators fully encumbered in current ALSE. This is somewhat counterintuitive in that investigations of this kind usually seek comparisons between “unloaded” and “loaded” levels of an independent variable. That human performance is degraded via the employment of MOPP ensembles and ALSE has been amply demonstrated and documented. It is, by far, more appropriate to estimate changes in performance in reference to current equipment and doctrine.

The approach to developing performance-based measurement scales is grounded in a framework of behavior-anchored ratings (Cain Smith & Kendall, 1963) of mission flights conducted in a simulator. This technique has been successfully and recently employed as a means of evaluating aircrews upon completion of aircrew coordination training, required of all Army aviators (U.S. Army Aviation Center, 1992). This investigation employs behavior-anchored rating scales that are already in use. The narrative text comprising the behavioral anchors has been modified or edited to reflect the fact that the ALSE, worn by the aviators while performing their mission flight, is the object of the evaluation. The behavior-anchored rating techniques and procedures of evaluation were adapted to the Air Warrior metrics purpose. It is not known, however, if this kind of behavioral metrics estimation technique has been employed to evaluate the influence, merits, or effect of a physical system such as ALSE.

The parent crew coordination evaluation methodology was validated in test bed development in 1992 (Simon, Grubb, & Leedom, 1993). Reliabilities of the original behavioral items were reported as “exceptionally high” (Simon, Grubb, & Leedom, 1992).

For obvious reasons, this investigation was a laboratory type effort obliged to be done in a flight simulation facility. The trials, however, were conducted in the AH-64 combat mission simulator (CMS), which is a fielded system intended for training of tactical units flying AH-64 Apache helicopters. The AH-64CMS is located at the Western Area Aviation Training Site (WAATS), Marana, Arizona.

METHOD

Subjects

Subjects consisted of ten crews of pilot and copilot-gunner pairs, involving a total of 20

aviators. They were all male Army National Guard pilots holding current medical certification, both rated and current in the AH-64 Apache and having mission experience in flying or in simulation while wearing full ALSE ensembles. Participating aviators were from units in Arizona, Idaho, South Carolina, and Utah. Apache aircrews are normally considered partnerships and do much of their training together. Except for one or two crews, paired crews were unavailable for this experiment; however, each aviator is trained to fly and fight from either seat, and they all had taken required crew coordination training within the preceding year. The rear seat pilot is usually the crew member on the flight controls and is the pilot in command. The front seat CPG aids in clearing the aircraft of obstacles, performs navigation, and normally acts as the sensors and weapons operator because of the intended panoramic outside view from this position. With very few exceptions, all operations may be conducted from either seat. Individual aviators begin their Apache flying in the front seat as CPGs and as they gain in training and experience, matriculate to the rear seat as those positions become available. This is true for both officers and warrant officers.

While not defined as subjects, the other participants in the investigation were the evaluators or scorers of the trials flown using the behavior-anchored rating scales. These were certified (U.S. Army Aviation Center, 1992, p. 3-1) crew coordination training evaluators. Evaluators may be flight examiners, instructor pilots, and unit trainers. There was one of each kind participating, all senior Warrant Officers, W-5, W-4, and W-3, respectively. All the evaluators were qualified AH-64CMS operators. Unfortunately, only one of these was able to participate in the entire number of trials, as shall be seen later in this report.

APPARATUS

Simulation Facilities

The flight simulator used in this investigation was the AH-64CMS, the combat mission simulator for the AH-64 Apache. It is normally employed as a tactical training, recurrency, and check ride platform for unit aircrews in lieu of actual aircraft. A single simulator combines two six-degree-of-freedom motion platforms (see Figure 1). The AH-64 rear seat cockpit is installed on one motion platform and the front seat on the other. This allows independent training in each cockpit or integrated training for crews and combat missions. The latter configuration was applied to this investigation.

Instrumentation

The AH-64CMS is capable of recording both event data (weapons selection, firing, hits

or misses, ownship hits, ground contact, etc.) and flight data (heading, ground track, altitude, airspeed, etc.) to be used as measures of effectiveness and quality of each flight trial.

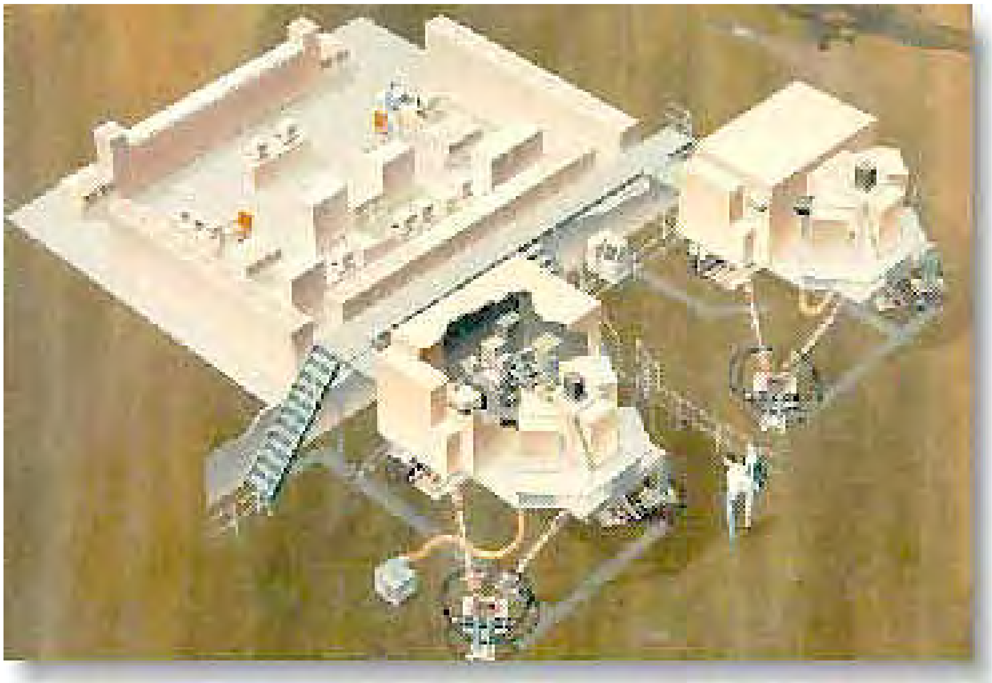


Figure 1. AH-64 combat mission simulator.

ALSE

Appendix A contains a listing of the equipment components to be worn by the subjects.

Procedures

SME Survey of Aviator and Aircrew Tasks

Given that here is a “strong correlation” between aircrew task performance and crew coordination behaviors (Simon, 1991), the presence of ALSE might be expected to have an impact on the aircrew’s performance of these tasks. Because 88 aviator and aircrew tasks are listed and described in the ATM for the AH-64, it made sense to attempt to discover if some tasks might be influenced to a greater degree than others via a survey taken among regular Army senior Apache aviators. If this were true, a mission scenario could embody those tasks most likely to be affected. Also, a scenario requiring that all 88 tasks be performed would impractical to the point of impossibility. The number of tasks selected for the survey was culled to 35 by several test pilot aviators at the U.S. Army Aviation Technical Test Center, Ft. Rucker, Alabama,

eliminating inappropriate tasks, such as flight planning (done outside the simulator), and redundant tasks, such as a variety of instrument approaches, as well as tasks that cannot be performed in the simulator. A sampling of some 12 other regular Army senior AH-64 aviators was surveyed to determine the relative influence that ALSE has on selected tasks by scoring both the task and its parts or subtasks and rank ordering the set of tasks placed in each of the scoring categories (very much affected, moderately affected, or slightly affected). Figure 2 illustrates the fact that there was no evidence of agreement among the SMEs as to which tasks might be affected by ALSE. Appendix B is a listing of the 35 surveyed ATM tasks in median ranked order.

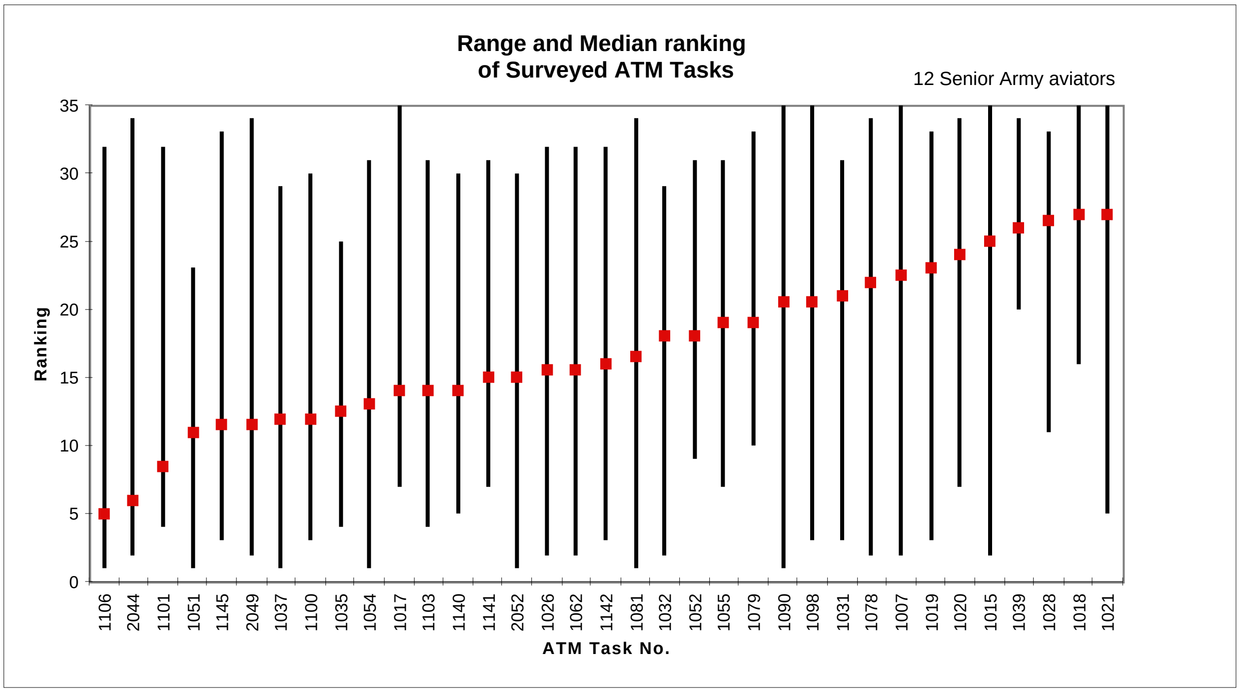


Figure 2. SME survey results.

The extremely wide range of SME rankings for the listed ATM tasks made the data unusable for building mission scenarios. However, scenario guidance and samples are provided in the Crew Coordination Evaluation Exportable Package (U.S. Army Aviation Center, 1992, Sect. 5).

Scenario

The Army training community, headed by the Directorate of Training and Simulation, Ft. Rucker, Alabama, has developed a standardized set of mission scenarios for the AH-64CMS, perhaps better described as tactical situations. Enemy and friendly forces are

arrayed on modeled and mapped terrain, and battle missions are devised for one or more Apaches in a flight, ownship and simulated other ships. Standardized published operational plans, operational orders, and pre-mission briefs are employed; however, the mission flight within this framework is completely open ended. This allowed the simulator operator-instructor, for example, to apply an engine failure or other system failure after the crew disengaged from the battle area in order to bring emergency procedure tasks into play. The standardized scenario employed for each crew's familiarization was named "Raid on Regimental Artillery Group" while the trials for record were conducted for "Hasty Attack Into Horse".

Methodology

The employment of behavior-anchored rating scales in the evaluation of aircrew performance as affected by ALSE, was accomplished by adapting the behavioral anchors (i.e., the text employed in the crew coordination evaluation methodology) to place the burden on behaviors as impacted by the equipment as opposed to pure coordination behaviors. All active duty Army and National Guard aviators are required to complete crew coordination training. Behavioral anchors were written for each of 13 basic qualities against which the evaluations are scored. Each basic quality contains three anchors with attendant text: Score value 1 (lowest), "very poor rating"; Score value 4 (midrange), "acceptable rating"; and Score value 7 (highest), "superior rating". Each of the original crew coordination anchors (U.S. Army Aviation Center, 1992, Sect. 6) has been edited or modified to reflect the effect of ALSE, as behavior overcame or succumbed to its influence. Appendix B contains an overview statement, a sampling of the modified behavioral anchors, plus a copy of the rating form (grade slip) for aircrew coordination evaluations, which was used without modification. Rating scores collected on grade slips, one for each crew member, for each trial, were to be the primary data for the investigation. Each trial also yielded measures of effectiveness data and flight data to be used in the analysis.

Grading and Scoring

The Army's standard simulator check ride procedure calls for evaluation and grading of each individual crew member against ATM tasks <u>and</u> crew coordination evaluation grading, that is, a single grade slip for the aircrew as an entity (not as individual aviators), and three grade slips in all. Because the wearing of ALSE may have different impacts on the individual aviators occupying the front and rear seats, the evaluators were required to execute an aircrew coordination evaluation grade slip for each individual but not to score and fill in the two grade slips for ATM tasks. The resulting workload for evaluators was thought to be approximately the same as with the standard check ride.

Trials Procedure

Subjects were first be briefed about the purpose of the investigation and procedures they were to follow and were allowed to ask questions of the investigators. It was made clear that the subjects were not on trial so much as was the current ensemble of ALSE components. The subjects were asked to do the best job that they could, however. They then read and signed the volunteer agreement. The standard procedure for launching a tactical mission was followed, including the dissemination of operations orders (OPORD) (see Appendix C), presentation of the air mission brief, scripted for the instructor and keyed to the scenarios and terrain residing in the simulator and pre-mission, flight and post-mission phases. Typically, each session lasted approximately 3-1/4 hours as follows: pre-mission planning and donning of ALSE, 1 hour; flight, 1-1/4 hours; post-mission, 1 hour.

Stress Assessment of Aviator Subjects[2]

The psychological and physiological state of soldiers while they performed combat-relevant tasks during stressful conditions may be critical to the outcome of a successful mission. The data presented here were designed to assess the psychological and physiological stress levels of aviators as affected by ALSE. A standardized battery of psychological state questionnaires was administered to the subjects in conjunction with a noninvasive physiological stress measure, salivary amylase. Both types of measures have been used in previous ARL research and have proved to be sensitive to the degree of stress experienced in a variety of situations.

Psychological State Measures

A 5-minute battery of stress perception measures was administered at strategic time points before, immediately following, and 1 hour after the each trial session. The following state measures were included:

Multiple Affect Adjective Check List

Revised Today Form (MAACL-R; Zuckerman & Lubin, 1985). The MAACL-R Today form consists of the following subscales: Anxiety, Depression, Hostility, Positive Affect. The scores are derived from a one-page list of 132 adjectives in which soldiers check all words that describe how they feel during a specified time period. This is used to

---

[2]The treatment of stress assessment information in the following sections was prepared by the following members of the HRED Soldier Stress and Cognitive Performance Team: Fatkin, Patton, Mullins, and Burton.

examine changes in specific affects in response to stressful situations. In addition, an overall distress score Dysphoria or Negative Affect was calculated by combining the Anxiety, Depression, and Hostility scores.

Specific Rating of Events scale (SRE)

The SRE requires the subjects to rate (on a scale of 0 to 100) how stressed they feel at the present moment or how they felt during a specified time period.

Subjective Stress Scale

This scale is designed to detect significant affective changes in stressful conditions (Kerle & Bialek, 1958). Soldiers are instructed to select one word from a list of 15 adjectives that best describes how they feel at the present moment or how they felt during a specified time period.

Physiological Assessment

Salivary amylase data were collected before, immediately following, and 1 hour after each trial session, including the familiarization procedure. These times coincide with the completion of the psychological state measures.

Amylase is an enzyme that hydrolyzes starch to oligosaccharides and then slowly to maltose and glucose. Measurement of amylase in saliva involves chemical color changes according to standard photometric procedures developed by Northwestern University (Chatterton, Vogelsong, Lu, Ellman, & Hudgens, 1996). This method combines time lapse and temperature data to derive a quantifiable level of stress.

Saliva samples for amylase assay were obtained from the subjects by using small, clean rectangular sponges (1 in. by 0.5 in. by 0.5 in.) in small plastic cups. The subjects were instructed to roll the sponges in their mouths for 1 minute while they completed the state questionnaires; then, upon instruction, they were to place the sponges back in the cups, snap the cover on them, and hand them to the investigator. The pre-labeled cups containing the sponges were then put in an insulated bag until the assays could be conducted.

Subject Training

The subjects were qualified and current in the AH-64 and were familiar with performing training exercises, practice missions, tactics, and taking check rides in an AH-64CMS (as well as in the AH-64 Apache). Annual recurrent training in MOPP IV is normally

required of all Army aviators, but the requirement is often waived. Additional ALSE components such as ballistic protection plates, life rafts, helicopter emergency egress device (HEED) (under water) air bottles, sidearms, and so forth are not normally worn during NBC training; therefore, few of the subjects had experienced flying or simulation missions dressed in the complete collection of encumbering components. A familiarization trial with the "Raid on Regimental Artillery Group" scenario was conducted before all trials for record.

Evaluator Training

Instructor pilots (IPs) and unit trainers (UTs) certified and serving as evaluators were trained in the specific procedures and scenarios associated with this investigation. Review and understanding of the modified basic qualities behavioral anchor narratives was emphasized. The evaluators practiced scoring trials during the first familiarization trials.

Experimental Design

The design for the conduct of trials was a Treatments x Subjects with repeated measures layout (Lindquist, 1953), where two repeated measures trials with the subjects fully encumbered were interleaved with one trial, per aircrew, with the over-water mission aspect eliminated, the variation trial. Ten AH-64 aircrews, that is, 20 aviators, participated. The individual members of an aircrew occupied the seat, front or rear, that they ordinarily had for the majority of their AH-64 experience. Rear seat subjects were designated S_1 through S_{10}, and the front seat subjects designated S_{11} though S_{20}. The series of three trials took place after the first familiarization trial had been completed. Table 1 illustrates the counterbalancing scheme[3] .

RESULTS[4]

Behavior-Anchored Ratings

Evaluators

Evaluators were a very rare commodity when it came to recruiting them for these trials. Evaluator No. 1, the CW5 Flight Examiner, was required to travel to Germany after only the first week of trials (Crews 1 through 3). He missed two individual trials during the time he was present. Evaluator No. 2 , the CW3 Unit Trainer, was present for the entire time trials were conducted. Both No. 1 and 2, however, missed the very first trial involving Crew No. 1.

---

[3]The counterbalancing scheme is not complete. Twelve crews would have provided a complete and balanced design, but only ten crews were available.
[4]An alpha level of .05 was used for all statistical tests.

Evaluator No. 3, the CW4 Instructor Pilot, was recruited midway through the trials (Crew 7 through 10), except he missed one of the trials for Crew No. 10. Table 2 summarizes the trials scored by each of the evaluators.

Table 1

Trials Order of Presentation by Crew and Subject No.

Crew No.	Trial 1	Trial 2	Variation trial
1 (S_1, S_{11})	1	2	3
2 (S_2, S_{12})	3	1	2
3 (S_3, S_{13})	2	3	1
4 (S_4, S_{14})	1	2	3
5 (S_5, S_{15})	3	1	2
6 (S_6, S_{16})	2	3	1
7 (S_7, S_{17})	1	2	3
8 (S_8, S_{18})	3	1	2
9 (S_9, S_{19})	2	3	1
10 (S_{10}, S_{20})	1	2	3

Table 2

Trials Scored by Participating Evaluators

	Trial 1			Trial 2			Variation trial		
Evaluator No. Crew No.	1	2	3	1	2	3	1	2	3
1 (S_1, S_{11})				X	X		X	X	
2 (S_2, S_{12})	X	X		X	X	X		X	
3 (S_3, S_{13})	X	X		X	X		X	X	
4 (S_4, S_{14})		X			X			X	
5 (S_5, S_{15})		X			X			X	
6 (S_6, S_{16})		X			X			X	
7 (S_7, S_{17})		X	X		X	X		X	X
8 (S_8, S_{18})		X	X		X	X		X	X
9 (S_9, S_{19})		X	X		X	X		X	X
10 (S_{10}, S_{20})		X	X		X			X	X

The evaluation scores on a Likert 7-point scale for the 13 basic qualities, common to Evaluator No. 1 and 2, were subjected to a hierarchical log linear (frequency) analysis. No statistically significant one-way effects for evaluators nor two-way associations ($p < .05$) were found. Scoring by Evaluator 1 and Evaluator 2 was not different.

The evaluation scores on a Likert 7-point scale for the 13 basic qualities, common to Evaluator No. 2 and 3, were subjected to a hierarchical log linear (frequency) analysis. No statistically significant one-way effects for evaluators nor two-way associations ($p < .05$) were found. Scoring by Evaluator 2 and Evaluator 3 was not different.

Crews

The evaluation scores on a Likert 7-point scale for the 13 basic qualities, collapsed or aggregated[5] for all evaluators, were subjected to a hierarchical log linear (frequency) analysis. Crew was the factor analyzed. No statistically significant one-way effects for individual crews nor two-way associations ($p < .05$) were found. Individual crew scores were not different.

Crew Position (pilot, rear seat versus copilot-gunner, front seat)

The evaluation scores on a Likert 7-point scale for the 13 basic qualities, aggregated for all evaluators, were subjected to a hierarchical log linear (frequency) analysis. Crew position was the factor analyzed. No statistically significant one-way effects for crew position nor two-way associations ($p < .05$) were found. Pilot and copilot-gunner scores were not different.

Trial Sequence

Given the counterbalanced design to accommodate the variation trial (no over-water equipment), the sequence of the three trials for record for each crew was examined for learning or adjustment effects. The evaluation scores on a Likert 7-point scale for the 13 basic qualities, aggregated for all evaluators, were subjected to a hierarchical log linear (frequency) analysis. Trial sequence was the factor analyzed. No statistically significant one-way effects for trial sequence nor two-way associations ($p < .05$) were found. Trial sequence scores were not different.

---

[5]Wherever it was considered appropriate in the analysis of data, the evaluation scores rendered by all the evaluators were employed. The first three crews and the last four crews had two sets of evaluation scores, recorded by two evaluators, for most of their trials.

Trial Type

The evaluation scores on a Likert 7-point scale for the 13 basic qualities, aggregated for all evaluators, were subjected to a hierarchical log linear (frequency) analysis. Trial type was the factor analyzed. No statistically significant one-way effects for trial sequence were found; however, two-way associations were noted for Basic Qualities 5, 7, 11 ($p < .05$). A further two-way association was noted for Basic Quality 12 when the first and second fully encumbered trial types only were analyzed. Overall, scores for trial type were not different.

Flying Experience

It was thought that the flying experience, in terms of both total flight hours or flight hours in the Apache, logged by the subjects was a potential factor for influencing the evaluation scores. The participating subjects represented a rather wide range of flying experience as seen in Figure 3.

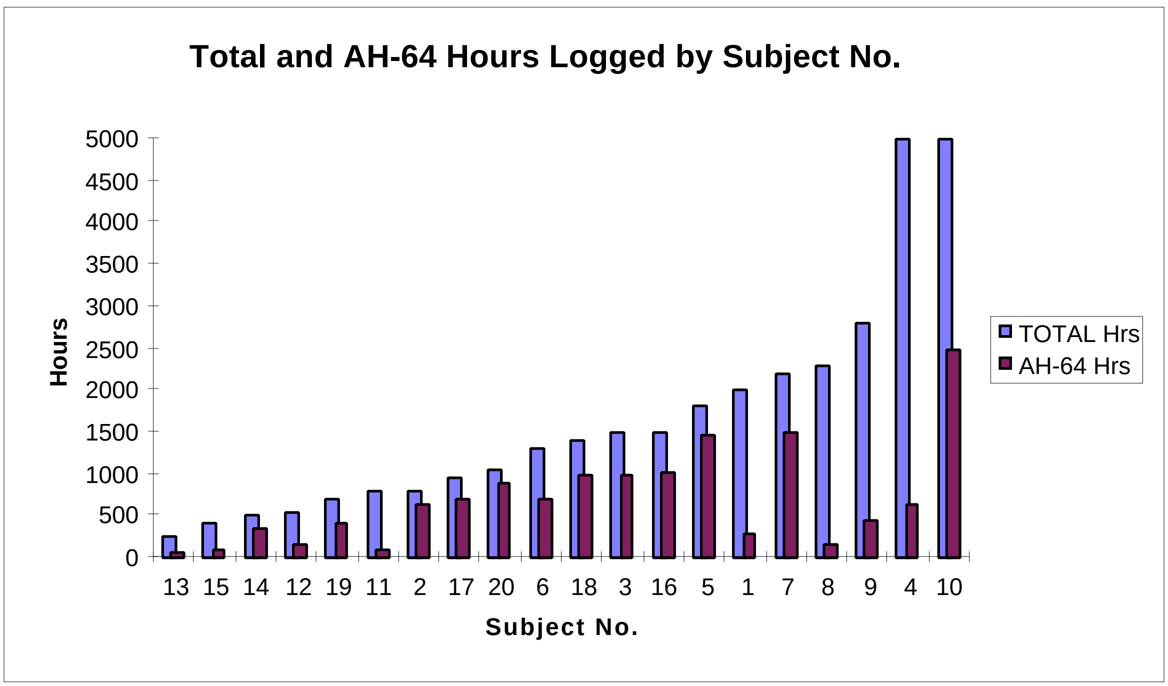


Figure 3. Flying experience of participating subjects.

The subjects were placed into the following experience categories (see Table 3) in order to facilitate a log linear analysis of the experience factor. There are 4 ± 1 subjects in each category.

The evaluation scores on a Likert 7-point scale for the 13 basic qualities, aggregated for all evaluators, were subjected to a hierarchical log linear (frequency) analysis. Trial sequence and trial type by total flight experience categories were the factors analyzed. No statistically significant one-way effects for trial sequence and trial type were found; however, some two-way and three-way associations were noted ($p < .05$). Total flight experience is not a factor with respect to trial sequence or trial type.

Table 3

Flying Experience Categories

Total flight categories	Total hours	AH-64 flight category	AH-64 hours
1	1 to 500	1	1 to 250
2	501 to 1000	2	251 to 500
3	1001 to 1500	3	501 to 750
4	1501 to 2000	4	751 to 1000
5	2001 to 5000	5	1001 to 2500

The evaluation scores on a Likert 7-point scale for the 13 basic qualities, aggregated for all evaluators, were subjected to a hierarchical log linear (frequency) analysis. Trial sequence and trial type by AH-64 flight experience categories were the factors analyzed. No statistically significant one-way effects for trial sequence and trial type were found; however, some two-way and three-way associations were noted ($p < .05$). AH-64 flight experience is not a factor with respect to trial sequence or trial type.

Measures of Effectiveness and Flight Data

Data in a number of different formats were collected during the conduct of trials in the AH-64CMS, reflecting the performance of the subjects in their tasks. Data such as the task times performed by the CPG in bringing the various Apache systems on line before takeoff were recorded by observation. Other measures were extracted from the simulator mainframe during the mission via hard copy requested via computer print command by the simulator operators-instructors-evaluators. In the press of conducting the trials, not all these data sets were recorded or recovered for each and every trial. At times, the simulator mainframe (1970’s architecture) was rather temperamental, and data were lost from working memory before they could be recovered. Main power fluctuations and maintenance problems in the computer room

environmental control units exacerbated the situation. The following analyses were performed with knowledge that the number of cases examined does not represent a complete data set for each and every trial.

CPG Task Times

The AH-64 crew's check list called for the CPG to initiate and check the operation of the navigation, sensor suite, fire control and weapons systems before takeoff and immediately after the pilot has started the ship's auxiliary power unit and brought it on line. The major tasks are

- Program the digital entry keyboard (DEK) weapons and ammunition data
- Helmet: integrated helmet and display sight system (IHADSS) operational check
- Program Doppler navigation set with navigation way points
- Target acquisition designation sight (TADS) operational check
- TADS initialization
- Forward looking infrared (FLIR) initialization
- Weapons initialization and check
- Fuel check

Also, the time from when the subjects were strapped in the cockpit to time of takeoff was recorded. Figure 4 illustrates the mean times and their standard deviations for the recorded data.

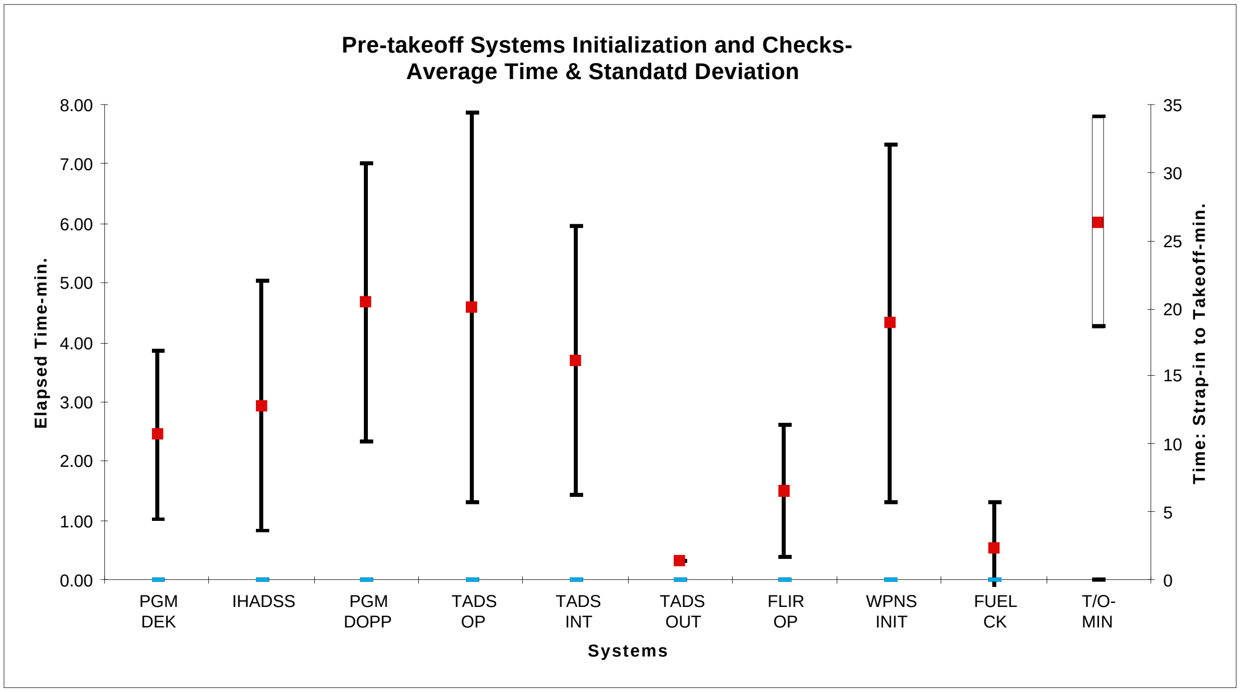


Figure 4. Mean times and standard deviations of CPG tasks.

Times for the data were subjected to one-way analyses of variance (ANOVAs) with CPG subjects as the between-groups effect. No statistically significant effect was found for any of these times with respect to trial sequence ($p < .05$). CPG subjects are considered a part of the same population.

Times for the data were subjected to one-way ANOVAs with trial sequence as the between-trials effect. No statistically significant effect was found for any of these times with respect to trial sequence ($p < .05$). The distributions of data points across trial sequence for these tasks were not necessarily even, possibly contributing to the lack of significance.

Times for the data were subjected to one-way ANOVAs with trial type as the between-trials effect. No statistically significant effect was found for any of these times with respect to trial type ($p < .05$). The distributions of data points across trial sequence for these tasks were not necessarily even, possibly contributing to the lack of significance.

Weapons Engagements

The Apache crews were ordered to fly from their starting point and conduct a hasty attack and destroy enemy forces in engagement area "HORSE," consisting of a tank battalion and a motorized rifle battalion. The enemy force provided opportunity for crews to use all their

weaponry, that is, HELLFIRE guided missiles, 2.75-inch rockets, and the 30-mm gun. Mission data were gathered for the total number of engagements, missile engagements, rocket engagements, and gun engagements; number of kills, hits, and misses, and the percentages of each; and the number of kills, hits, and misses with each type of weapon. Figure 5[6] summarizes the number of kills, hits, and misses for each type of weapon. Figure 6 summarizes the percentage of kills, hits, and misses and the total number of engagements for each weapon.

All these measures were subjected to one-way ANOVAs with crews as the between-groups effect. No statistically significant effect was found for any of these times with respect to trial sequence ($p < .05$). Crews are considered a part of the same population.

All these measures were subjected to one-way ANOVAs with trial sequence as the between-trials effect. No statistically significant effect was found for any of these times with respect to trial sequence ($p < .05$). There were no differences in any of these measures with respect to trial sequence.

[6]FirstR denotes the first trial for record fully encumbered, 2ndR is the second trial for record fully encumbered, VR is the variation (no over-water components) trial for record. These are the three trial types referenced.

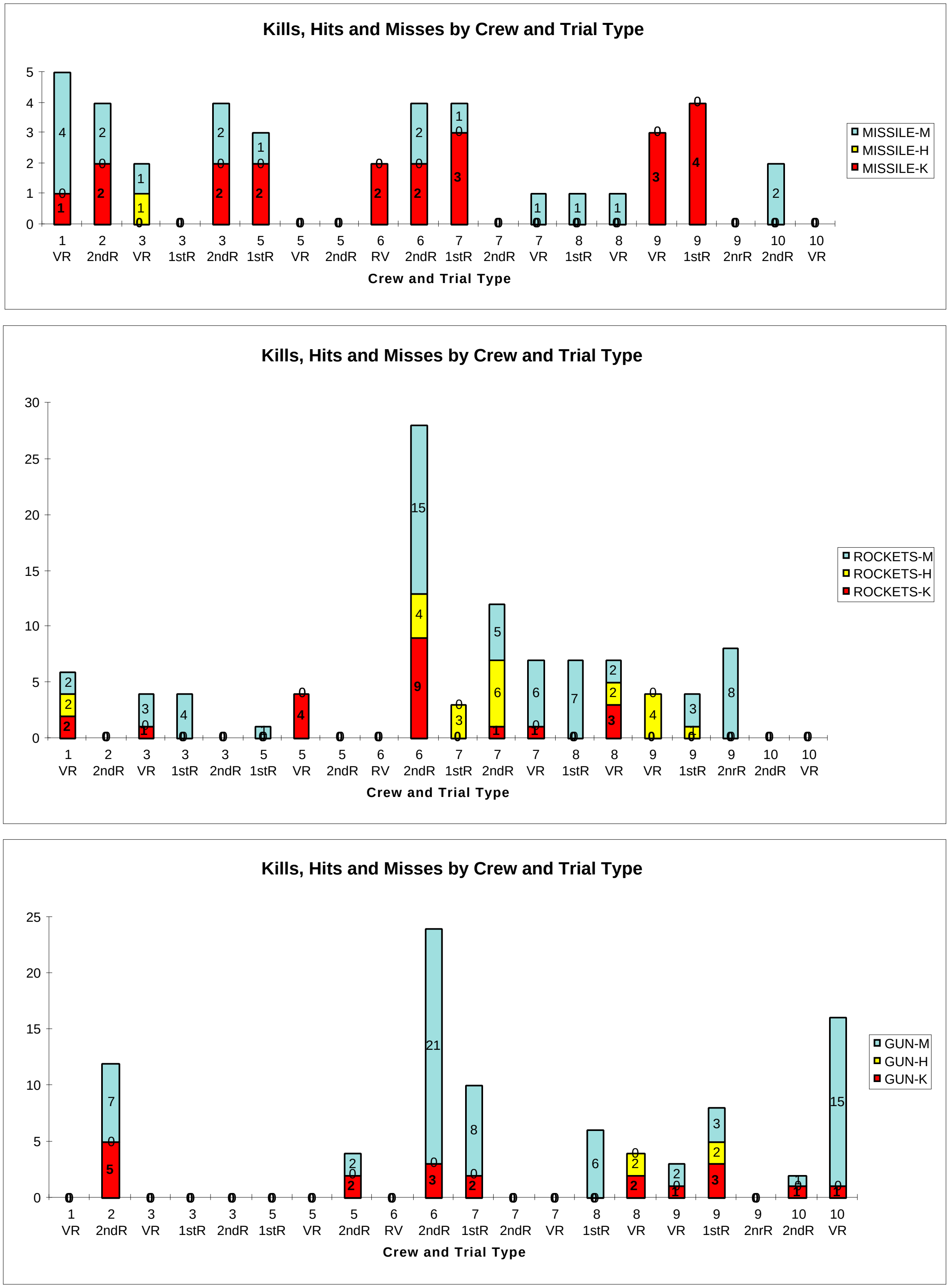


Figure 5. Kills, hits, and misses.

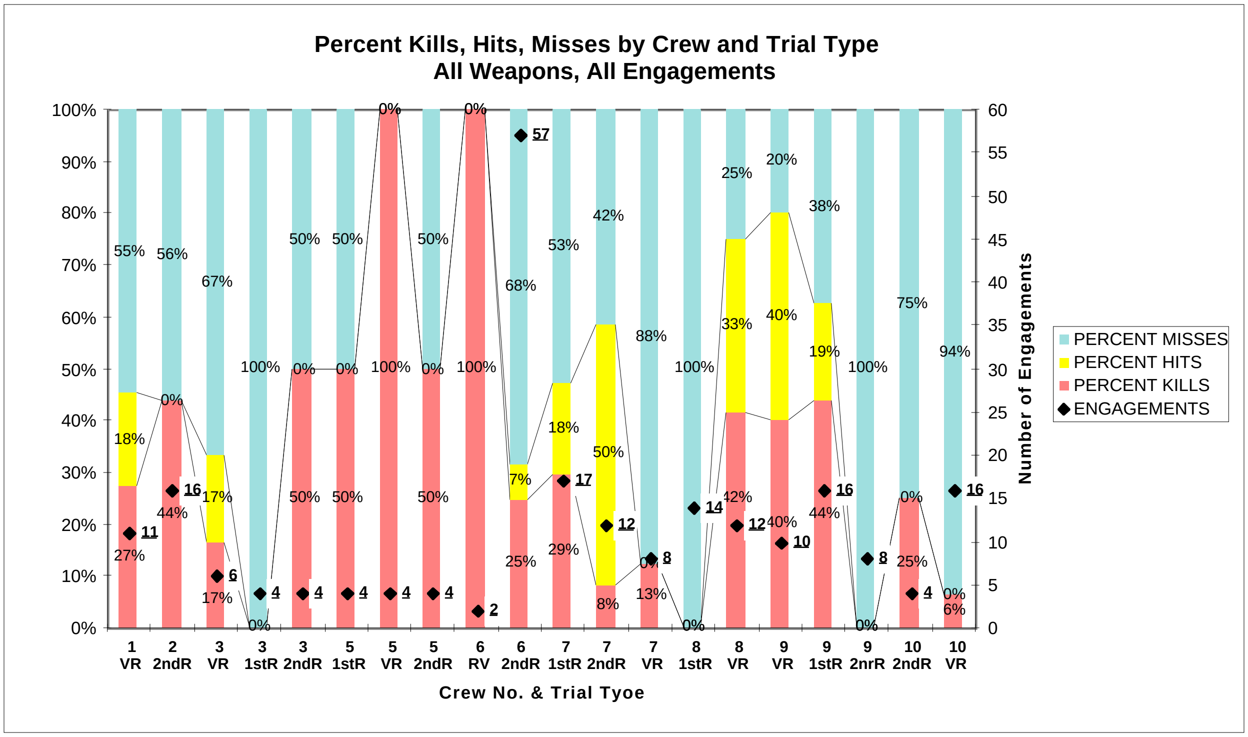


Figure 6. Percentage kills, hits, and misses.

All these measures were subjected to one-way ANOVAs with trial type as the between-trials effect. No statistically significant effect was found for any of these times with respect to trial type ($p < .05$). There were no differences in any of these measures with respect to trial type.

Weapons Platform Stability

The AH-64CMS momentarily records the Apache's motion during engagements in terms of pitch, yaw, and roll rates and vertical and lateral translation rates. All are in units of feet per second, plus (right and upward velocities) and minus (left and downward velocities) movement of the launcher rail or gun muzzle. These data are not available for missile rounds. This is done from 5 seconds before trigger pull until the last rocket or gun round has left the ship and constitutes an "engagement". For both rockets and gun, an engagement may be a burst of rounds with 5 seconds to the next trigger pull. Data were collected for the maximum velocities in the plus direction and the maximum velocities in the minus direction and algebraically summed, regardless of whether these maximums occurred in the pitch, yaw, roll, vertical or lateral axes. The resulting sums then represent a range of rates from plus to minus for further

analysis. It would have been more rigorous to perform a vector summation of the rate data; however, the results would have been more conservative and compacted.

The platform stability range of velocities for AH-64CMS recorded trials was subjected to one-way ANOVAs with crews as the between-trials effect. A statistically significant effect was found with respect to trial sequence ($p = .001$). Both Tukey and Duncan post hoc tests were applied to these data. They yielded slightly different results (see Table 4). Note, however, that the number of trials representing the simulator recorded data is 2 to 3, extremely small.

Table 4

Results of Both Tukey and Duncan Post Hoc Tests for Platform Stability—Crews

Crew No.	5	3	9	10	6	7
N	3	2	3	2	2	3
Means	1.030	1.040	1.063	1.095	1.135	1.153

Tukey

Duncan

$p < .05$

Means underscored by the same line are not significantly different.
Means not underscored by the same line are significantly different

The Tukey test says Crew No. 5, 3, 9, and 10 were not significantly different from each other, but grouped, they were different from Crew No. 6 and 7; Crew No. 9, 10, and 6 were not significantly different from each other, but grouped, they were different from Crew No. 3, 2, and 7 and Crew No. 6 and 7; Crew No. 10, 6, and 7 were not significantly different from each other, but grouped, they were different from Crew No. 5, 3, and 9. The Duncan test says Crew No. 5, 3, and 9 were not significantly different from each other, but grouped, they were different from Crew No. 10, 6, and 7; Crew No. 9 and 10 were not significantly different from each other,

but grouped, they were different from Crew No. 5, 3, 6, and 7; Crew No. 10 and 6 were not significantly different from each other, but grouped, they were different from Crew No. 5, 3, 9, and 7; and Crew No. 6 and 7 were not significantly different from each other, but grouped, they were different from Crew No. 5, 3, 9, and 10.

The platform stability range of velocities for AH-64CMS recorded trials was subjected to one-way ANOVAs with trial sequence as the between-trials effect (see Table 5). No statistically significant effect was found for any of these times with respect to trial sequence (p < .05). There were no differences in any of these measures with respect to trial sequence. The sample sizes for the trial means are small.

Table 5

Mean Platform Stability Range of Velocities—Trial Sequence

Trial sequence	1	2	3
N	5	6	6
Means	1.072	1.070	1.117

The platform stability range of velocities for AH-64CMS recorded trials was subjected to one-way ANOVAs with trial type as the between-trials effect (see Table 6). No statistically significant effect was found for any of these times with respect to trial type (p < .05). There were no differences in any of these measures with respect to trial type. Again, the sample sizes for the trial means are small.

Table 6

Mean Platform Stability Range of Velocities—Trial Type

Trial type	First R	Second R	VR
N	4	6	7
Means	1.058	1.100	1.093

Enemy Forces Threat to Ownship

The enemy forces deployed in EA HORSE could shoot back at the Apache crews during their mission and did so if and when the crew's ownship was searched, found, and acquired for a long enough period. While the enemy's tanks and troop carriers and command vehicles could engage with guns, the ownship was out of their range more often than not. The bulk of the threat was from a variety of air defense weapons accompanying the forces, which employed guns and rockets and missiles, optics and radar. The probabilities of acquisition and hit, given ownship shot at is a function of how much of the ownship is unmasked (by exposure zone), the time of exposure, time of flight of the enemy projectiles, and whether the ownship can remask. Figure 7 describes the ownship exposure zones portion of the probability functions.

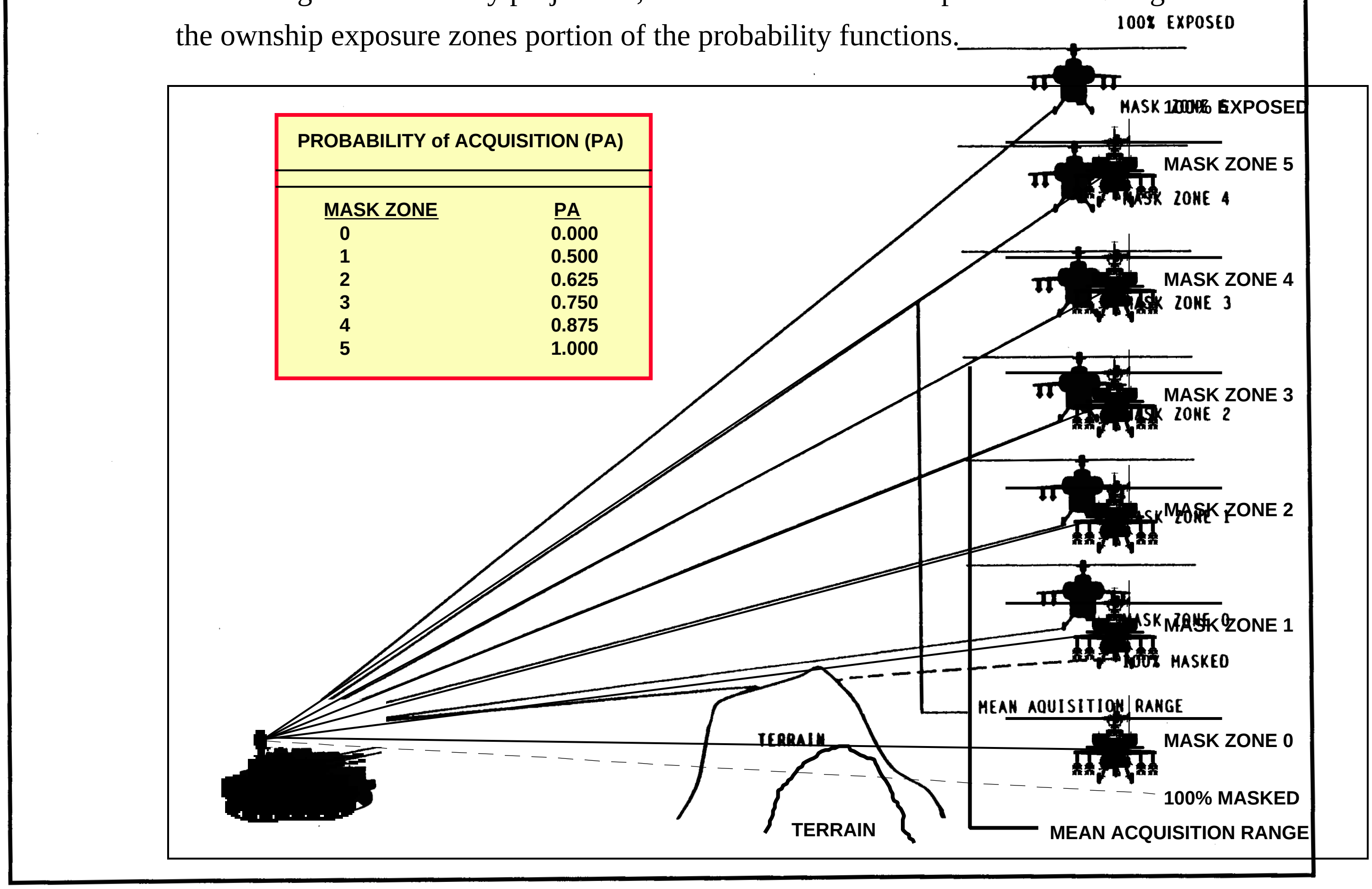


Figure 7. Ownship exposure zones.

The AH-64CMS computes a threat "event" with its attendant acquisition and hit probabilities whenever the ownship is exposed, with line of sight in exposure zone one or higher for 10 seconds or more, with 5 seconds or more continuous (U.S. Army Materiel Command, Simulation, Training, and Instrumentation Command, 1996). The number of events occurring on a given mission is uncontrolled and solely dependent on the pilot's flying maneuvers. Measures

of summed maximum exposure zones (over all the events), of maximum probabilities of enemy acquisition, maximum probabilities of enemy hits, total exposure time, average exposure time per event, percent of events where shot at, hits, misses, and no-shots were examined. Figures 8, 9, and 10 show the threat data gathered.

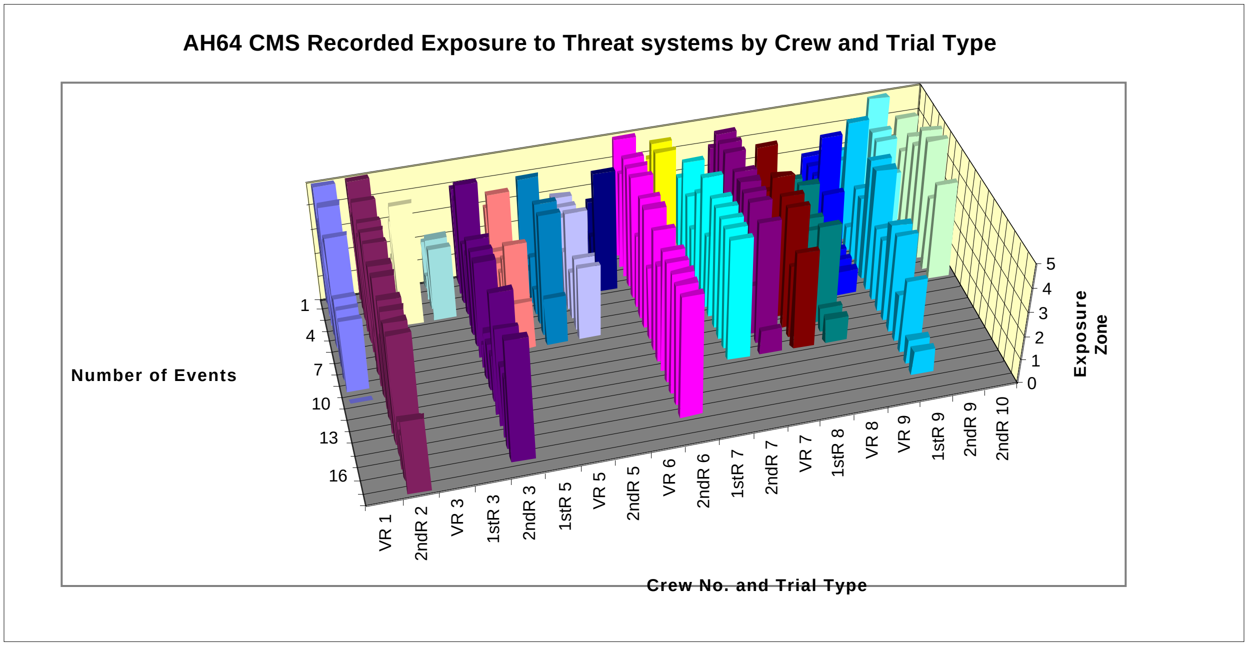


Figure 8. Events: Maximum exposure zones.

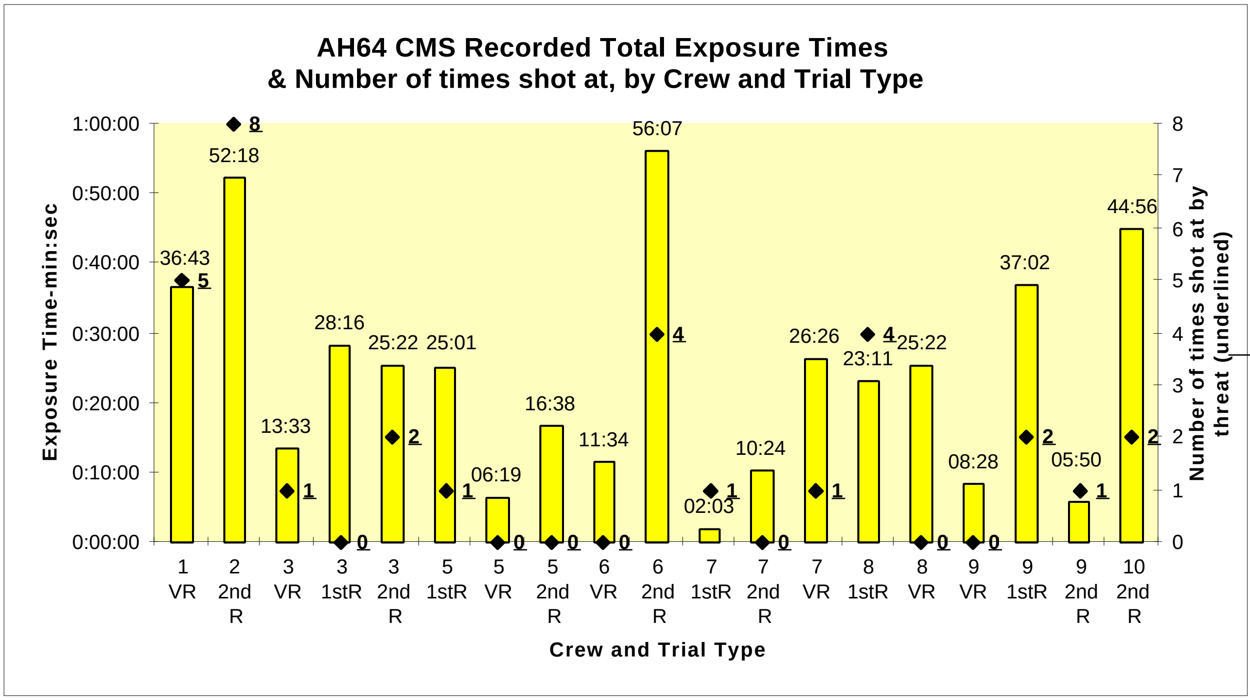


Figure 9. Total exposure times and times shot at.

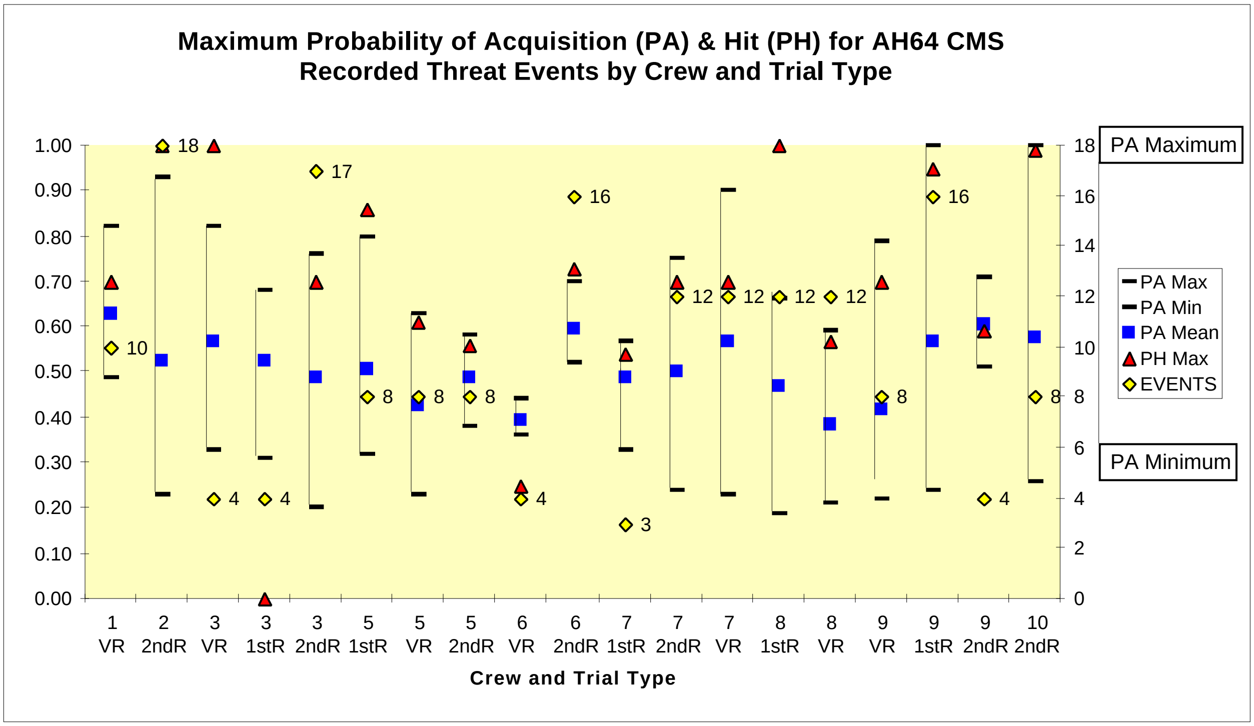


Figure 10. Acquisition and hit probabilities.

The sum of maximum exposure zones for AH-64CMS recorded trials was subjected to one-way ANOVAs with trial sequence as the between-trials effect. A statistically significant effect was found with respect to trial sequence (p = .048). Both Tukey and Duncan post hoc tests were applied to the data. Both yielded the same results (see Table 7).

Table 7

Results of Both Tukey and Duncan Post Hoc Tests for
Summed Exposure Zones—Trial Sequence

Trial sequence	1	2	3
N	6	6	7
Means	20.00	31.00	47.14
Tukey and Duncan	______________	______________	
		______________	______________
p < .05			

Means underscored by the same line are not significantly different.
Means not underscored by the same line are significantly different

The first trial for record was not significantly different from the second but was different from the third. The second trial for record was not significantly different from either of the other two.

The sum of maximum exposure zones for AH-64CMS recorded trials was subjected to one-way ANOVAs with trial type as the between-trials effect (see Table 8). No statistically significant effect was found for any of these times with respect to trial type ($p < .05$). There were no differences in any of these measures with respect to trial type.

Table 8

Mean Sums of Maximum Exposure Zones—Trial Type

Trial type	First R	Second R	VR
N	5	7	7
Means	27.20	46.71	24.71

The maximum probability of acquisition for AH-64CMS recorded trials was subjected to one-way ANOVAs with trial sequence as the between-trials effect. A statistically significant effect was found for with respect to trial sequence ($p = .017$). Both Tukey and Duncan post hoc tests were applied to the data. They yielded slightly different results (see Table 9).

Table 9

Results of Both Tukey and Duncan Post Hoc Tests for
Probability of Acquisition—Trial Sequence

Trial sequence	1	2	3
N	6	6	6
Means	.3517	.4317	.6233

Tukey

Duncan

$p < .05$

Means underscored by the same line are not significantly different.
Means not underscored by the same line are significantly different

The Tukey test says that the first trial for record was not significantly different from the second but was different from the third, and the second trial for record was not significantly different from either of the other two. The Duncan test says the first trial for record was not significantly different from the second, but the third trial for record was different from both the first and second.

The maximum probability of acquisition for AH-64CMS recorded trials was subjected to one-way ANOVAs with trial type as the between-trials effect (see Table 10). No statistically significant effect was found for any of these times with respect to trial type ($p < .05$). There were no differences in maximum probability of acquisition measures with respect to trial type.

Table 10

Mean Maximum Probabilities of Acquisition—Trial Type

Trial type	First R	Second R	VR
N	5	6	7
Means	.3700	.5300	.4871

The maximum probability of hit for AH-64CMS recorded trials was subjected to one-way ANOVAs with trial sequence as the between-trials effect. A statistically significant effect was found for with respect to trial sequence ($p = .002$). Both Tukey and Duncan post hoc tests were applied to the data. They yielded the same results (see Table 11).

Table 11

Results of Both Tukey and Duncan Post Hoc Tests for
Maximum Probability of Hit—Trial Sequence

Trial sequence	1	2	3
N	6	6	6
Means	20.00	31.00	47.14

Tukey and Duncan ________________________________

$p < .05$

Means underscored by the same line are not significantly different.
Means not underscored by the same line are significantly different

Both tests note that the first trial for record was not significantly different from the second, but the third trial for record was different from both the first and second.

The maximum probability of hit for AH-64CMS recorded trials was subjected to one-way ANOVAs with trial type as the between-trials effect. A statistically significant effect was found for with respect to trial type ($p = .031$). Both Tukey and Duncan post hoc tests were applied to the data. They yielded the slightly different results (see Table 12).

Table 12

Results of Both Tukey and Duncan Post Hoc Tests for
Maximum Probability of Hit—Trial Type

Trial type	First R	Second R	VR
N	5	6	7
Means	.0000	.2464	.3257
Tukey	__________	__________	
		__________	__________
Duncan		__________	__________
$p < .05$			

Means underscored by the same line are not significantly different.
Means not underscored by the same line are significantly different

The Tukey test says the first trial for record was not significantly different from the second but was different from the VR trials, and the second trial for record was not significantly different from either of the other two. The Duncan test says the first trial for record was significantly different from the second and the VR trials. Note that for the FirstR trials, the hit probabilities were all zeros, but that is for only five crews.

The data (number of events, ownship hits, ownship misses, no shot at ownship, percent shot at, total exposure time, average exposure time per event, sum of exposure zones for AH-64CMS recorded trials) were subjected to one-way ANOVAs with both trial sequence and trial type as the between-trials effect. Means are shown in Tables 13 and 14. No statistically significant effect was found for any of these measures with respect to trial sequence or trial type ($p < .05$). There were no differences in any of these measures with respect to trial sequence or trial type.

Table 13

Means for Number of Events, Ownship Hits, Ownship Misses, No Shot at Ownship, Percent Shot At, Total Exposure Time, Average Exposure Time per Event, Sum of Exposure Zones for Trial Sequence

			Number of events	Ownship hits	Ownship misses	No shot at ownship	Percent shot at	Total exposure time	Average exposure time per event	Sum of exposure zones
Trial sequence	1	Mean	6.50	.17	1.00	5.33	17.35	13:58	02:10	20.00
		N	6	6	6	6	6	6	6	6
		SD	3.45	.41	1.10	2.66	15.45	08:46	01:08	11.73
	2	Mean	10.00	.17	.50	9.33	6.25	25:23	03:07	31.00
		N	6	6	6	6	6	6	6	6
		SD	4.20	.41	.84	3.93	10.46	14:55	02:37	15.86
	3	Mean	12.14	.00	3.00	9.14	23.50	31:20	02:28	47.14
		N	7	7	7	7	7	7	7	7
		SD	5.18	.00	2.83	4.14	18.55	18:17	00:54	23.52
	Total	Mean	9.68	.11	1.58	8.00	16.11	23:58	02:35	33.47
		N	19	19	19	19	19	19	19	19
		SD	4.78	.32	2.12	3.93	16.29	15:48	01:38	20.68

SD = standard deviation

Table 14

Means for Number of Events, Ownship Hits, Ownship Misses, No Shot at Ownship, Percent Shot At, Total Exposure Time, Average, Exposure Time per Event, Sum of Exposure Zones, by Trial Type

			Number of events	Ownship hits	Ownship misses	No shot at ownship	Percent shot at	Total exposure time	Average exposure time per event	Sum of exposure zones
Trial type	firstR	Mean	8.60	.20	1.40	7.00	18.32	23:06	03:01	27.20
		N	5	5	5	5	5	5	5	5
		SD	5.46	.45	1.14	4.58	14.60	12:55	02:25	17.51
	secondR	Mean	11.86	.14	2.29	9.43	18.74	30:13	02:33	46.71
		N	7	7	7	7	7	7	7	7
		SD	5.37	.38	2.87	4.08	15.95	20:41	01:37	23.80
	VR	Mean	8.29	.00	1.00	7.29	11.90	18:20	02:17	24.71
		N	7	7	7	7	7	7	7	7
		SD	3.35	.00	1.83	3.45	19.16	11:16	01:06	13.61
	Total	Mean	9.68	.11	1.58	8.00	16.11	23:58	02:35	33.47
		N	19	19	19	19	19	19	19	19
		SD	4.78	.32	2.12	3.93	16.29	15:48	01:38	20.68

An intercorrelation matrix of the threat measures is shown in Table 15. Of interest is that Maximum Probability of Acquisition by Exposure Zone Maximum (r = .729) and Maximum Probability of Acquisition by Maximum Probability of Hit (r =.771) exceed the $r^2 \geq .50$ commonality rule of thumb.

Table 15

Threat Measures Intercorrelations

Pearson's ρ correlation	Ownship hits	Maximum probability of acquisition	Maximum probability of hit	Total exposure time	Average exposure time per event	Percent shot at
Ownship hits						
Maximum probability of acquisition	-.146					
Maximum proba-of hit	.057	.771**				
Total exposure time	.225	.159	.372			
Average exposure time per event	.256	.081	.085	.564*		
Percent shot at	.282	.213	.266	.424	.128	
Exposure zone maximum	-.151	.729**	.610**	.338	-.101	.449

* Correlation is significant at p<.05 (two-tailed).
** Correlation is significant at p<.01 (two-tailed).

Stress Assessment of Aviator Subjects

A repeated measures ANOVA was conducted on the state data with two factors: time (pre, post, recovery) and trial type (familiarization, first fully encumbered trial, second fully encumbered trial, and variation trial) with seat (rear, pilot versus front, CPG) included as a between-group variable. The combined MAACL-R data revealed a main effect of time, $F(2,18) = 11.25$, $p < .001$, with post measures being significantly higher than pre or recovery measures. In addition, there was a Subscore x Time interaction, $F(8,72) = 9.06$, $p < .001$. Table 16 and Figure 11 present the overall means, pre, post and recovery for the MAACL-R combined data. Table 17 and Figure 12

show the mean scores for the MAACL-R subscales and the Subscore x Time interaction. There was no effect of seat for the combined data.

Table 16

MAACL-R Today Combined Data Means

Means MAACL-R	Combined data
Pre	50.12
Post	61.67
Recovery	49.40

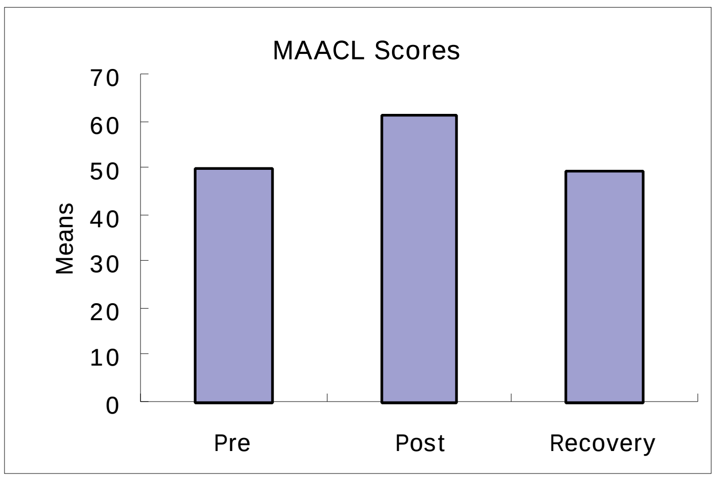


Figure 11. Histogram, MAACL-R today combined data means.

Table 17

Mean Scores for MAACL-R Subscales: Anxiety, Depression,
Hostility, Positive Affect, and Dysphoria

Means MAACL data	Pre	Post	Recovery
Anxiety	48.29	53.05	46.58
Depression	47.04	69.72	47.44
Hostility	49.15	67.68	48.70
Positive affect	58.44	51.06	57.78
Dysphoria	47.68	66.81	46.51

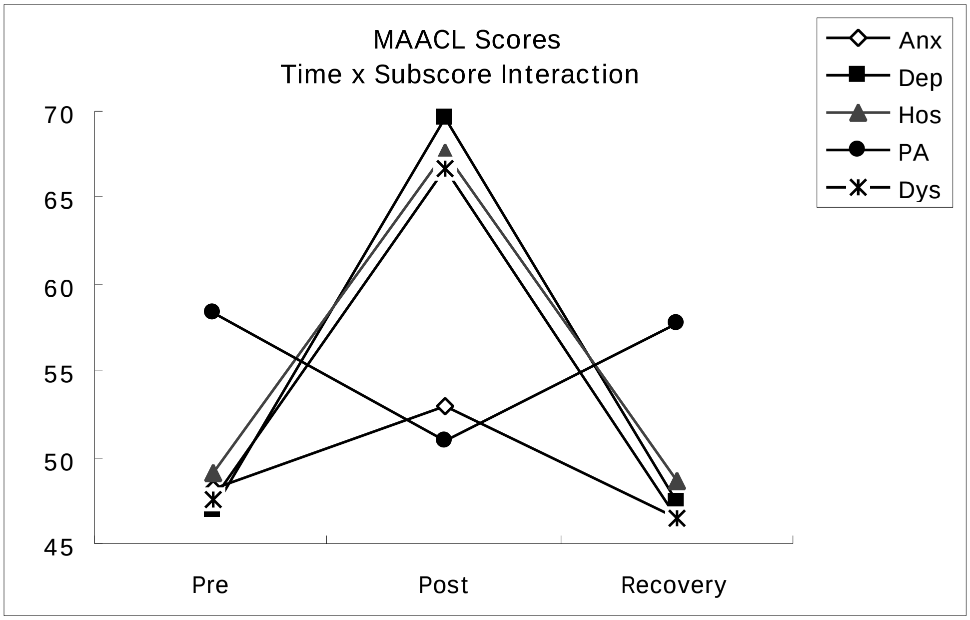


Figure 12. Mean scores for MAACL-R subscales interaction.

A separate repeated measures ANOVA was conducted for each of the MAACL-R subscores as well as the subjective stress data and the SRE data with two factors: time (pre, post, recovery) and trial type (familiarization, first fully encumbered trial, second fully encumbered trial, and variation trial) and with seat (rear, pilot versus front, CPG) included as a between-group variable. There was no main effect of trial or seat for any of the subscores.

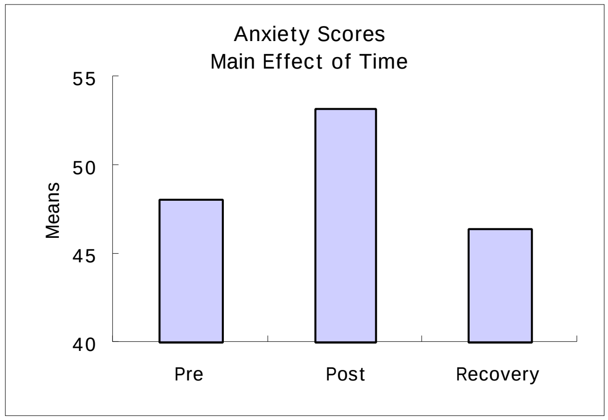
MAACL-R Anxiety

There was a significant main effect of time on the MAACL-R Anxiety subscale $\underline{F}(2,18) = 7.26$, $\underline{p} = .005$. The level of anxiety was higher immediately following a trial than before or 1 hour after a trial (see Table 18 and Figure 13).

Table 18

Mean Anxiety Scores by Time

Means	Anxiety scores
Pre	48.10
Post	53.21
Recovery	46.47



<u>Figure 13.</u> Histogram, mean anxiety scores by time.

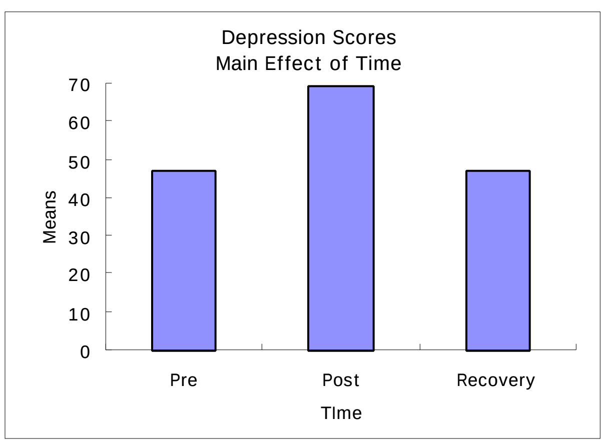
MAACL-R Depression

There was a significant main effect of time on the MAACL-R Depression subscale $\underline{F}(2,18) = 11.94$, $\underline{p} = .001$. The level of depression was higher immediately following a trial than before or 1 hour after a trial (see Table 19 and Figure 14). Furthermore, there was a significant Trial x Time interaction $\underline{F}(6,54) = 2.71$, $\underline{p} = .023$ (see Table 20 and Figure 15).

Table 19

Mean Depression Scores by Time

Means	Depression scores
Pre	47.04
Post	69.72
Recovery	47.44



<u>Figure 14.</u> Histogram, mean depression scores by time.

Table 20

Mean Depression Scores by Time and Trial

Means	Depression scores Pre	Post	Recovery
Trial familiarization	45.92	82.70	46.75
Trial firstR	48.75	65.00	47.00
Trial secondR	46.50	69.00	49.00
Trial VR	47.00	62.20	47.00

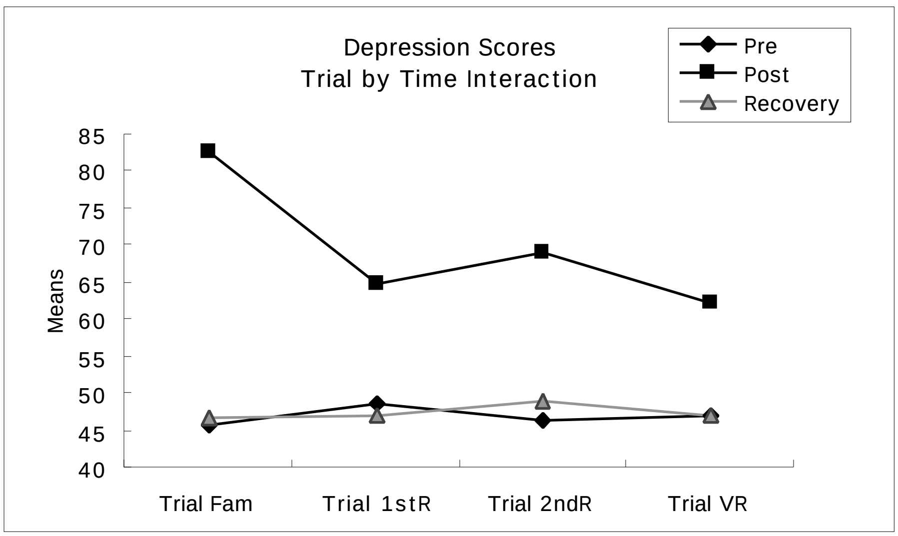


Figure 15. Mean depression scores by time and trial.

MAACL-R Hostility

There was a significant main effect of time on the Hostility subscale $F(2,18) = 7.53$, $p = .004$. The level of hostility was higher immediately following a trial than before or 1 hour after a trial (see Table 21 and Figure 16).

Table 21

Mean Hostility Scores by Time

Means	Hostility scores
Pre	49.37
Post	67.46
Recovery	48.65

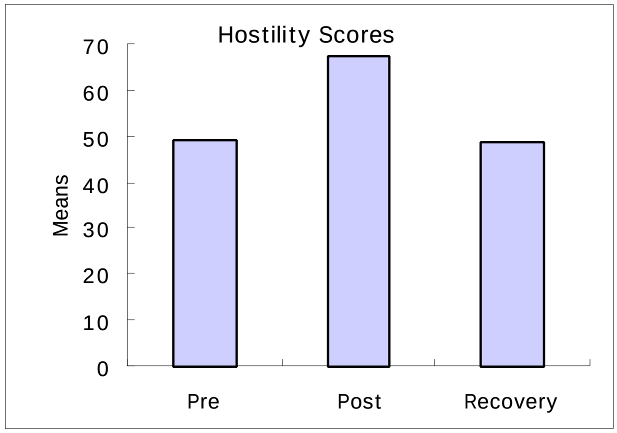


Figure 16. Histogram, mean hostility scores by time.

MAACL-R Dysphoria (Negative Affect)

There was a significant main effect of time on the Dysphoria or Negative Affect score $\underline{F}(2, 18) = 12.04$, $\underline{p} = .001$. The level of negative affect was higher immediately following a trial than before or 1 hour after a trial (see Table 22 and Figure 17).

Table 22

Mean Dysphoria Scores by Time

Means	Dysphoria scores
Pre	47.68
Post	66.81
Recovery	46.51

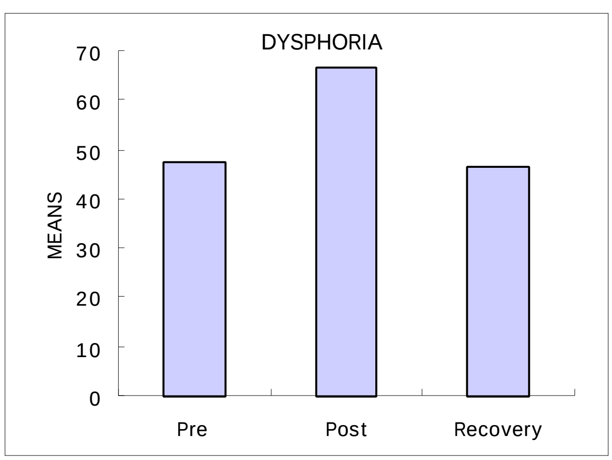


Figure 17. Histogram, mean dysphoria scores by time.

MAACL-R Positive Affect

There was a significant main effect of time on the Positive Affect subscale $\underline{F}(2, 18) = 9.03$, $\underline{p} = .002$. The level of Positive Affect was lower immediately following a trial than before or 1 hour after a trial (see Table 23 and Figure 18). There was also a significant Trial x Time interaction $\underline{F}(6, 54) = 2.59$, $\underline{p} = .028$ (see Table 24 and Figure 19).

Table 23

Mean Positive Affect Scores by Time

Means	Positive affect scores
Pre	58.6
Post	51.4
Recovery	57.9

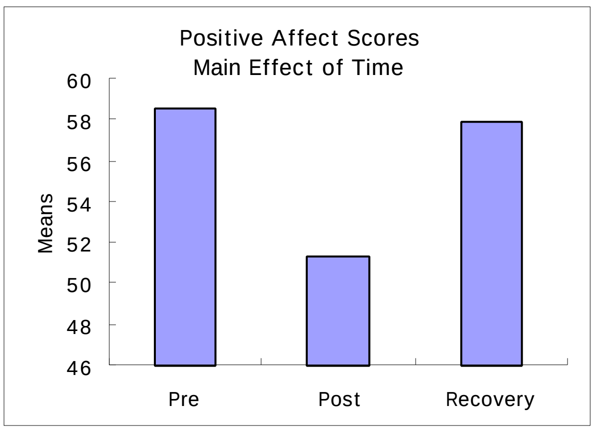


Figure 18. Histogram, mean positive affect scores by time.

Table 24

Mean Positive Affect Scores by Time and Trial

Means	Positive affect scores Pre	Post	Recovery
Trial familiarization	60.63	50.22	60.83
Trial firstR	54.90	54.33	57.48
Trial secondR	61.53	49.30	53.93
Trial VR	56.70	50.38	58.88

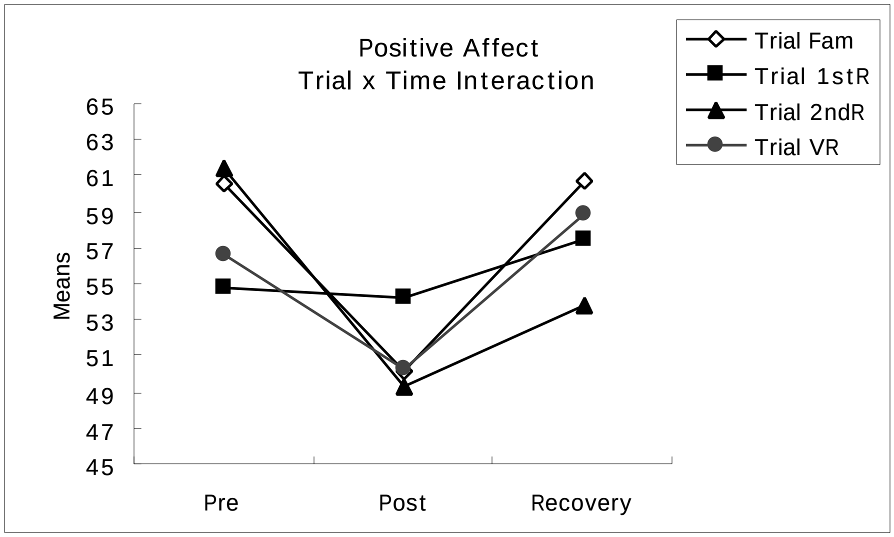


Figure 19. Mean positive affect scores by time and trial.

Subjective Stress Scale

There was a significant main effect of time on the Subjective Stress scale $F(2, 36) = 22.83$, $p = .000$. Subjects reported higher levels of stress immediately following a trial than before or 1 hour after a trial (see Table 25 and Figure 20). No significant main effects were found for either trial type or seat; however, a Trial x Seat interaction $F(3, 54) = 2.93$, $p = .042$ revealed that pilots in the rear seat reported higher stress for the first trial for record than did the CPGs in the front seat. Conversely, during the second trial, the rear seat pilots claimed less stress whereas the front seat CPGs reported an increase in subjective stress (see Table 26 and Figure 21). A three-way Seat x Trial x Time interaction was also noted $F(6, 108) = 3.41$, $p = .004$.

Table 25

Subjective Stress Scores by Time

Means	Subjective stress scores
Pre	22.24
Post	46.53
Recovery	18.76

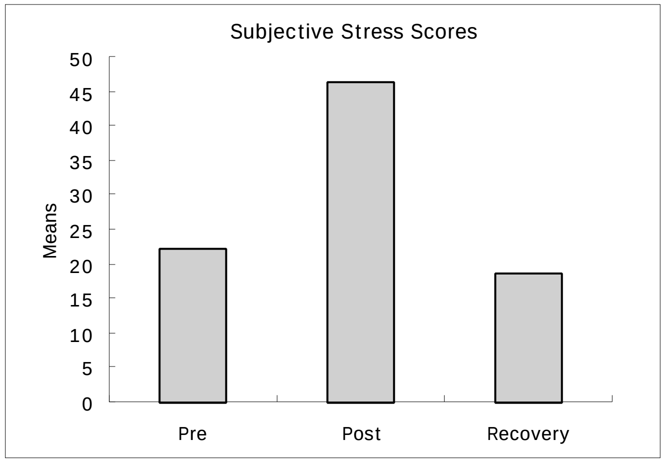


Figure 20. Histogram, subjective stress scores by time.

Table 26
Mean Subjective Stress Scores by Trial and Seat

Means	Subjective stress scores Front	Rear
Trial x Seat	Front	Rear
Trial familiarization	31.40	28.40
Trial firstR	23.93	36.83
Trial secondR	31.67	33.20
Trial VR	27.03	20.93

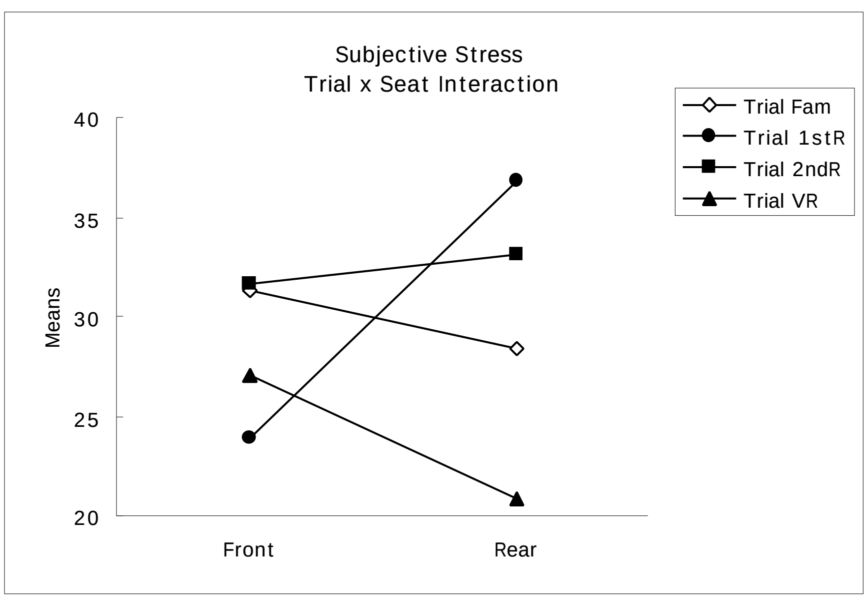


Figure 21. Mean positive affect scores by trial and seat.

Specific Rating of Events Scale (SRE)

There was a significant main effect of time on the overall stress levels as measured by the SRE scale $F(2, 36) = 53.34$, $p = .000$. Subjects reported higher overall stress immediately after a trial than they did before or 1 hour after a trial (see Table 27 and Figure 22).

There were no significant findings of main effects for seat or trail, nor two-way interactions for seat, trial or time. There was a three-way Seat x Trial x Time interaction $F(6, 108) = 2.36$, $p = .035$.

Table 27

SRE Scores by Time

Means	SRE scores
Pre	17.26
Post	55.66
Recovery	12.00

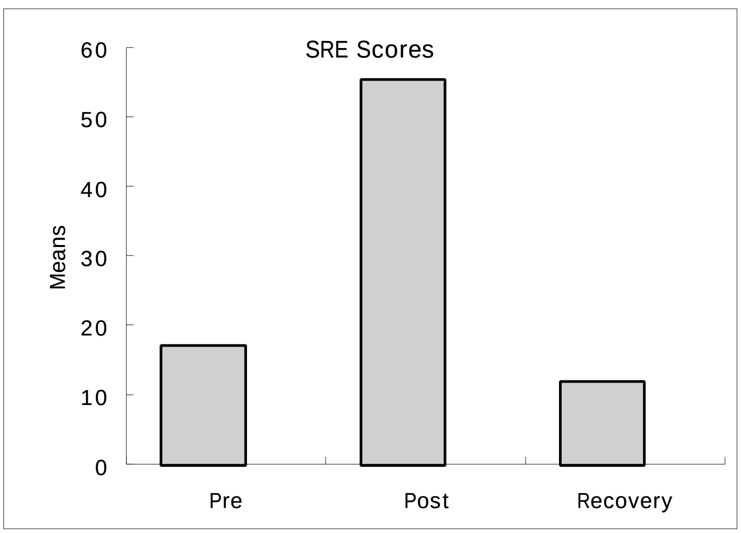


Figure 22. Histogram, SRE scores by time.

Salivary Amylase

A repeated measures ANOVA was conducted on the salivary amylase data with two factors: time (pre, post, recovery) and trial type (familiarization, first fully encumbered trial, second

fully encumbered trial, and variation trial) with seat (front versus back) included as a between-group variable (see Table 28 and Figure 23).

The effect of trial type did not reach significance nor were there any significant differences between seats for the amylase data. Only a main effect of time was noted $\underline{F}(2,20) = 3.60$, $\underline{p} = .046$. Amylase measures were significantly higher immediately following a trial than they were before or 1 hour after testing.

Table 28

Amylase Data Scores by Time

Means	Amylase data scores
Pre	230.77
Post	238.41
Recovery	226.75

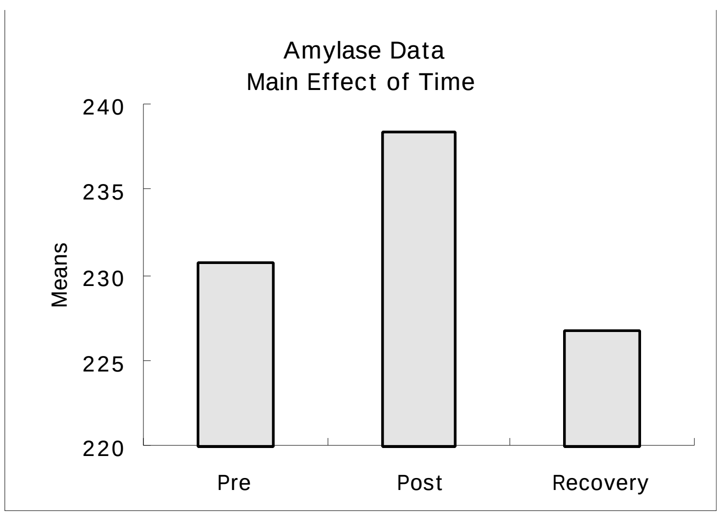


Figure 23. Histogram, amylase data scores by time.

Stress Assessment Batteries

The psychological state data were compared with data from five referent protocols (Torre et al., 1991; Fatkin & Hudgens, 1994; Tauson, Doss, Rice, Tyrol, & Davidson, 1995). Each of these referent protocols included a pre-stress measurement and a post-stress measurement. The five referent protocols for the present evaluation are

1. ONCOL SURG—men visiting a hospital on a day when their wives were facing cancer surgery under general anesthesia.

2. ABDOM SURG—men visiting a hospital on a day when their wives were facing abdominal surgery under general anesthesia,

3. WR EXAM—third year male medical students taking a written examination required for completion of the clerkship portion of their medical training.

4. SS COMPET—male soldiers representing elite units in marksmanship (sharp shooter) competition.

5. INDEP CONTROL—men investigated during normal work days when they were experiencing no unusual stress.

The ONCOL SURG and ABDOM SURG protocols represent a relatively high stress level when compared with the WR EXAM and SS COMP protocols, which represent a relatively moderate level of stress. The INDEP CONTROL protocol represents a relatively low stress level to a condition of no stress. The INDEP CONTROL pre-stress responses are an independent, unstressed measure, which was used as a basis for comparison against the current results (see Figures 24 through 36). Error bars in these figures are standard error of the means.

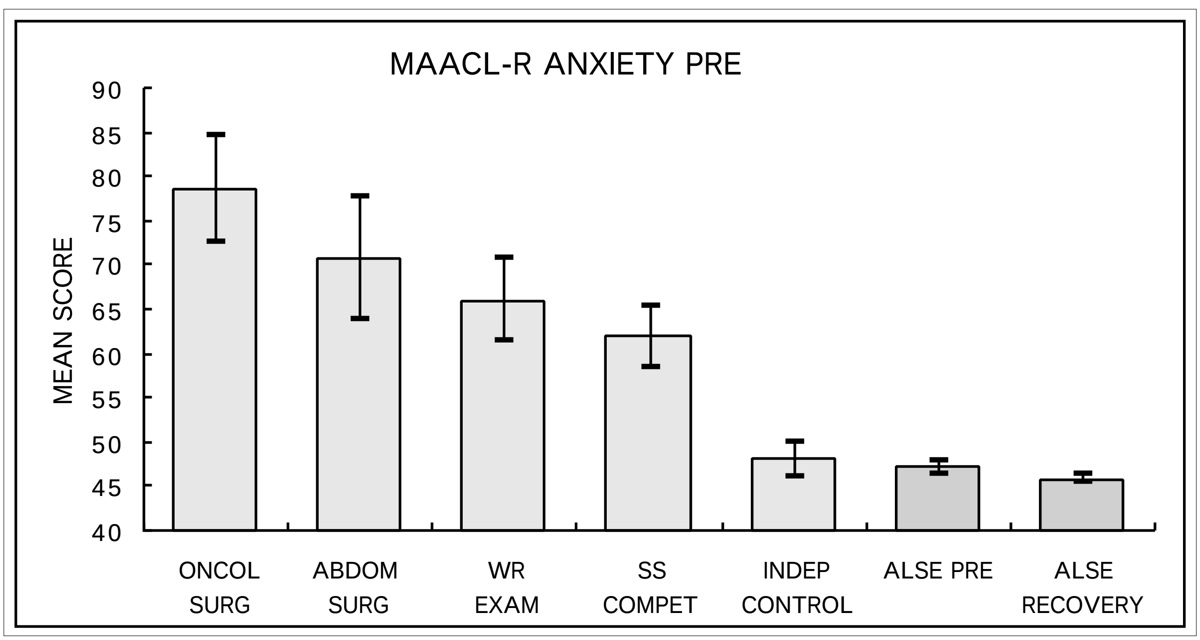


Figure 24. Comparison of mean pre-stress anxiety scores.

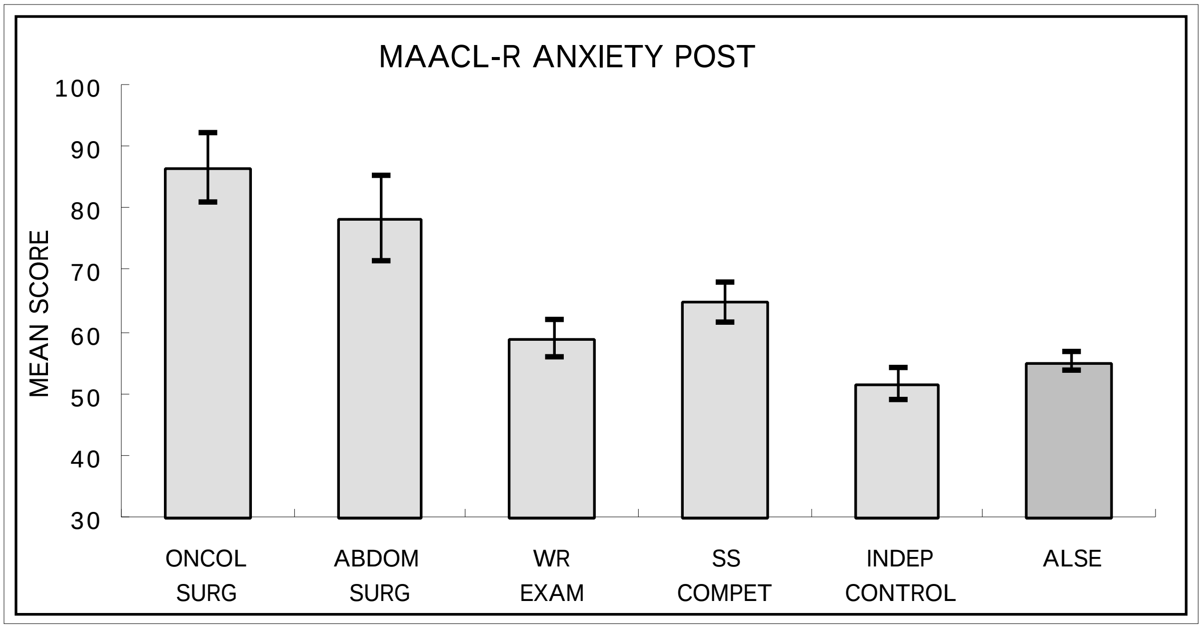


Figure 25. Comparison of mean post-stress anxiety scores.

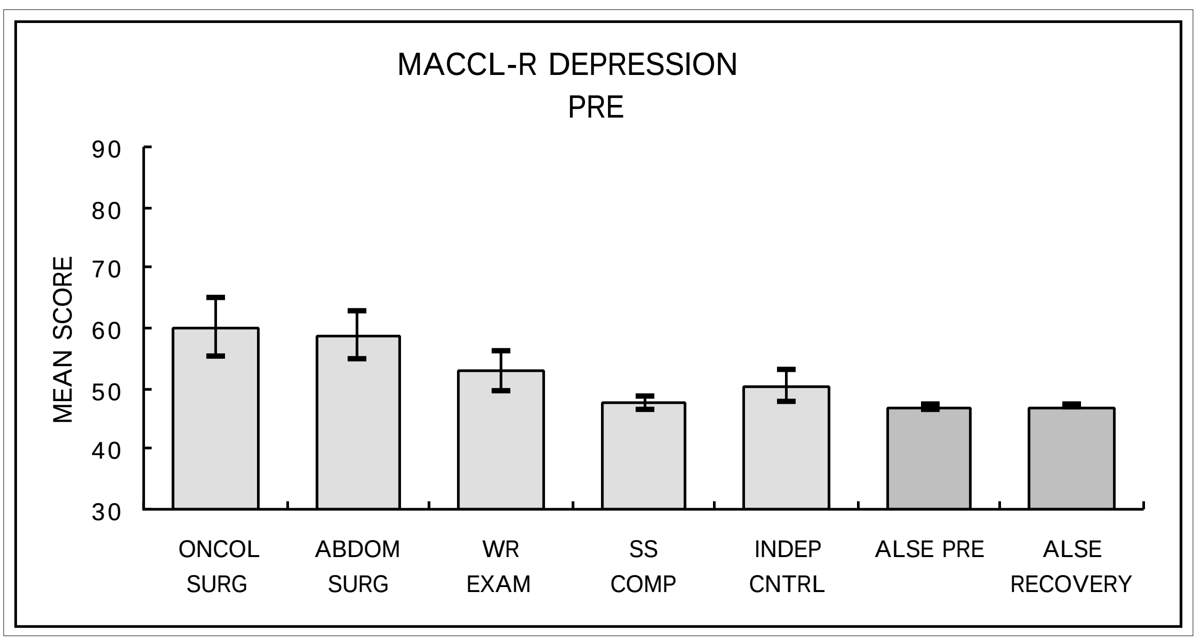


Figure 26. Comparison of mean pre-stress depression scores.

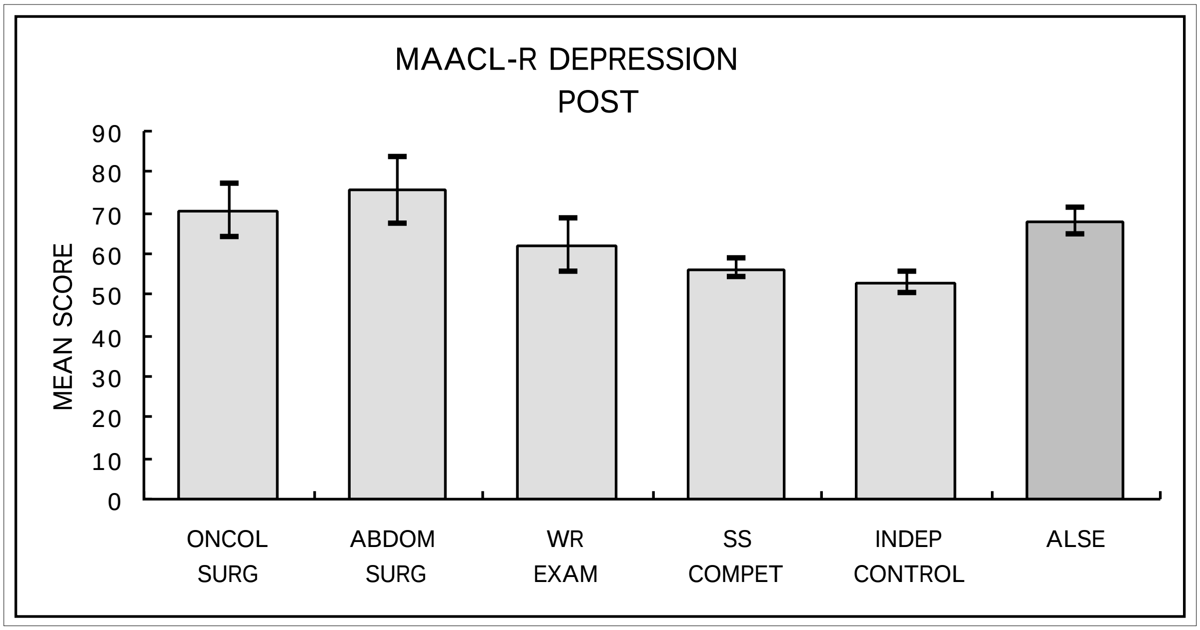


Figure 27. Comparison of mean post-stress depression scores.

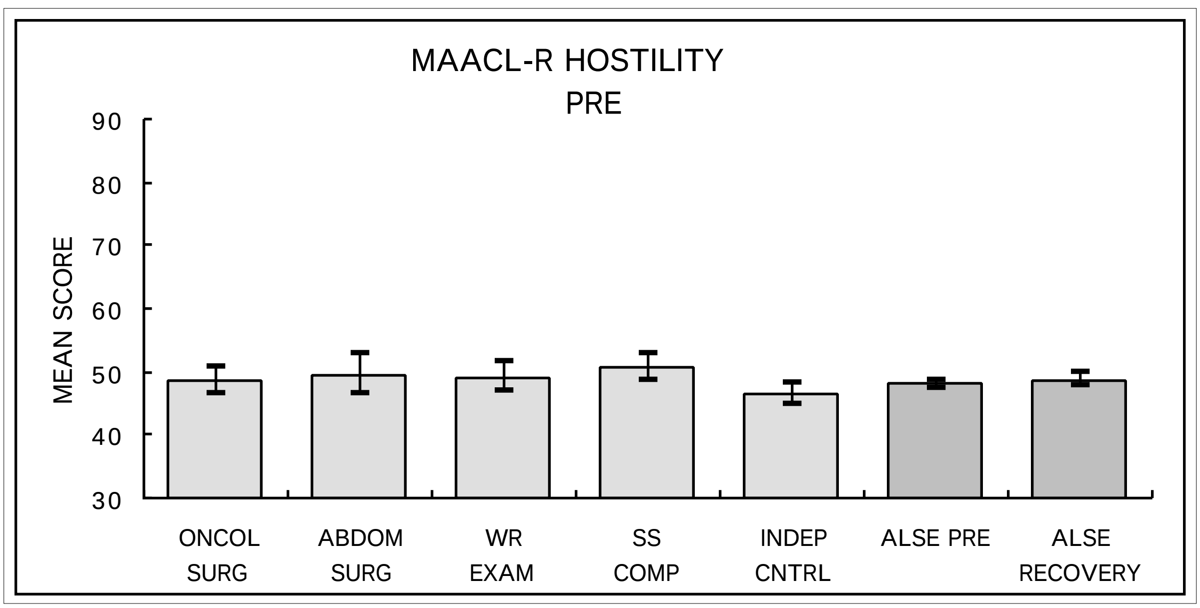


Figure 28. Comparison of mean pre-stress hostility scores.

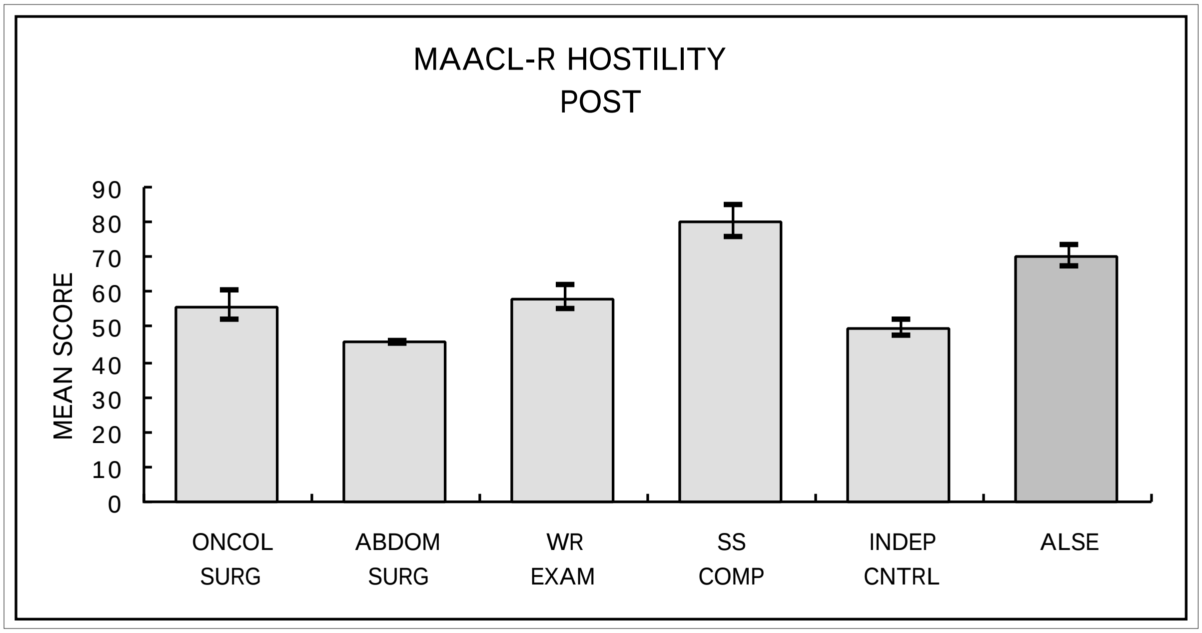


Figure 29. Comparison of mean post-stress hostility scores.

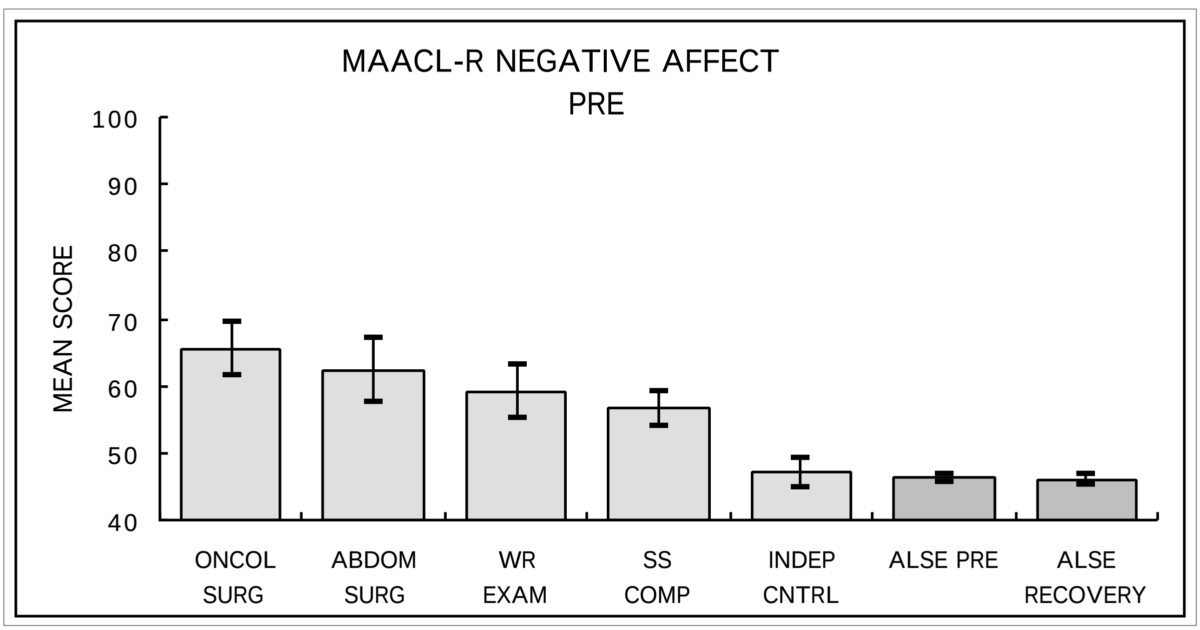


Figure 30. Comparison of mean pre-stress negative affect scores.

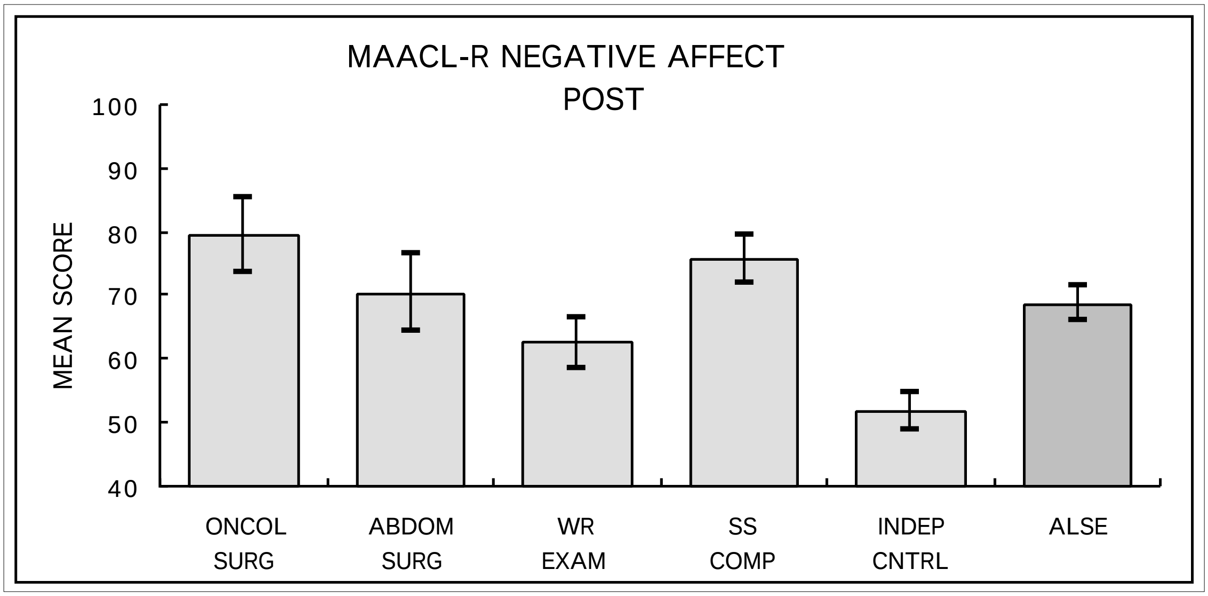


Figure 31. Comparison of mean post-stress negative affect scores.

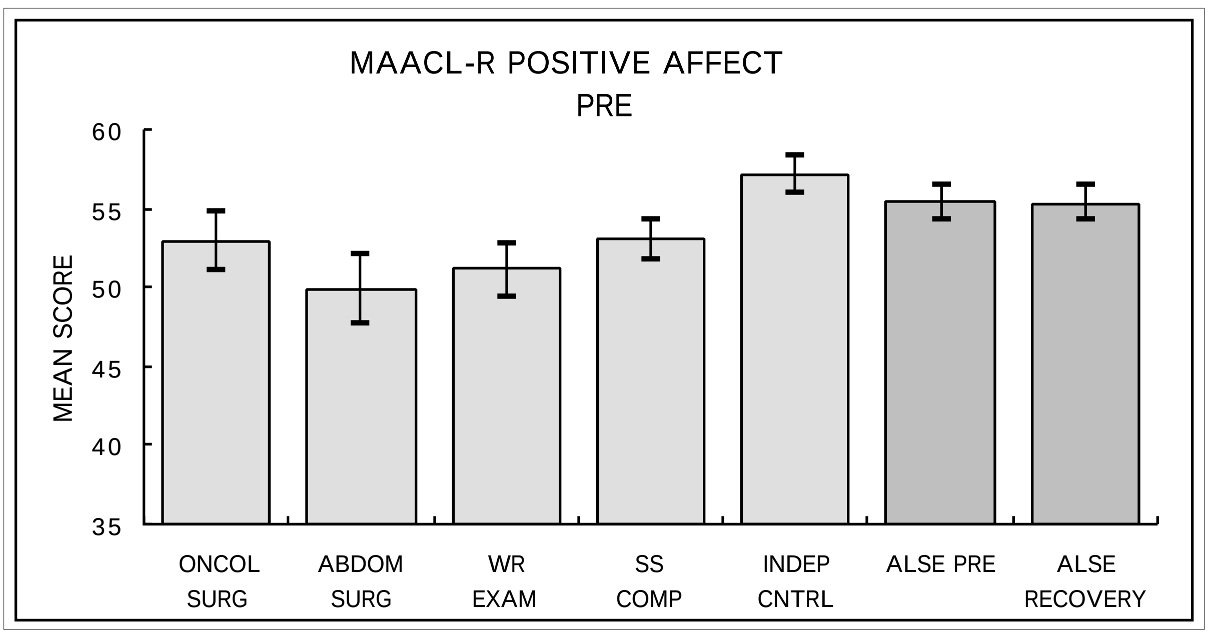


Figure 32. Comparison of mean pre-stress positive affect scores.

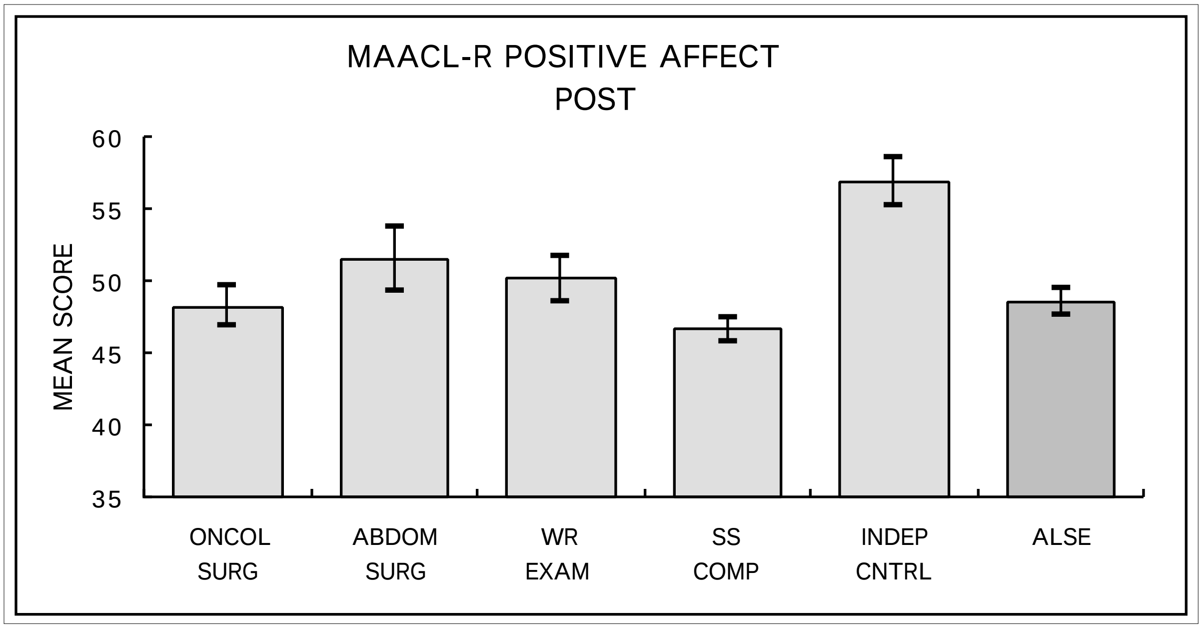


Figure 33. Comparison of mean post-stress positive affect scores.

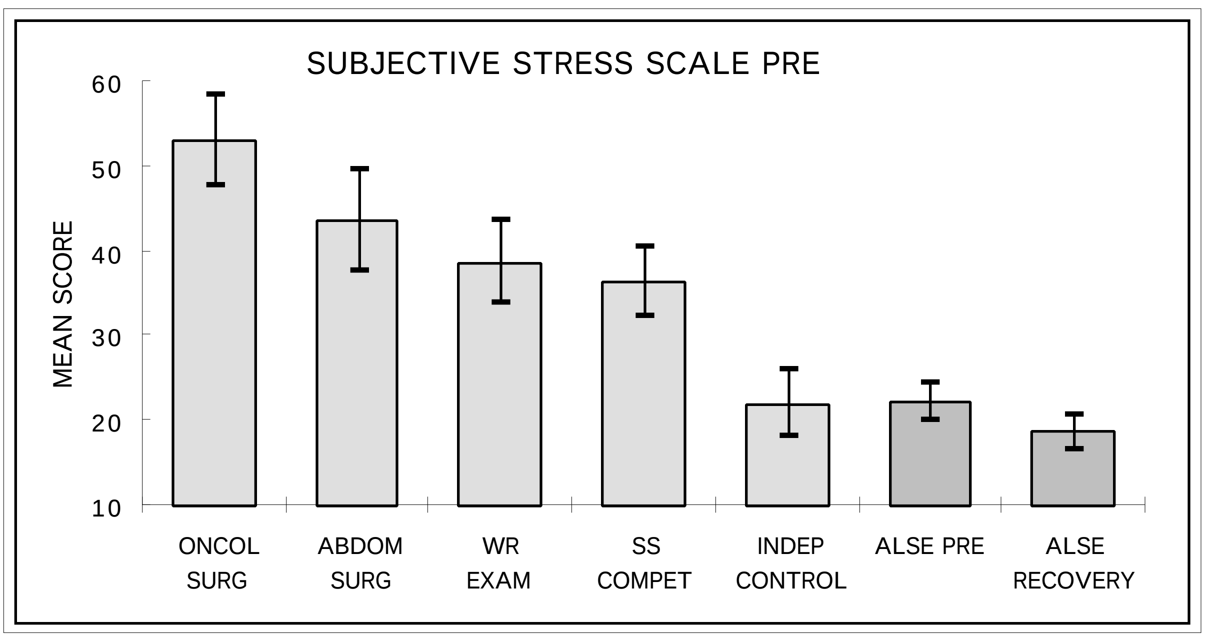


Figure 34. Comparison of mean pre-stress subjective stress scores.

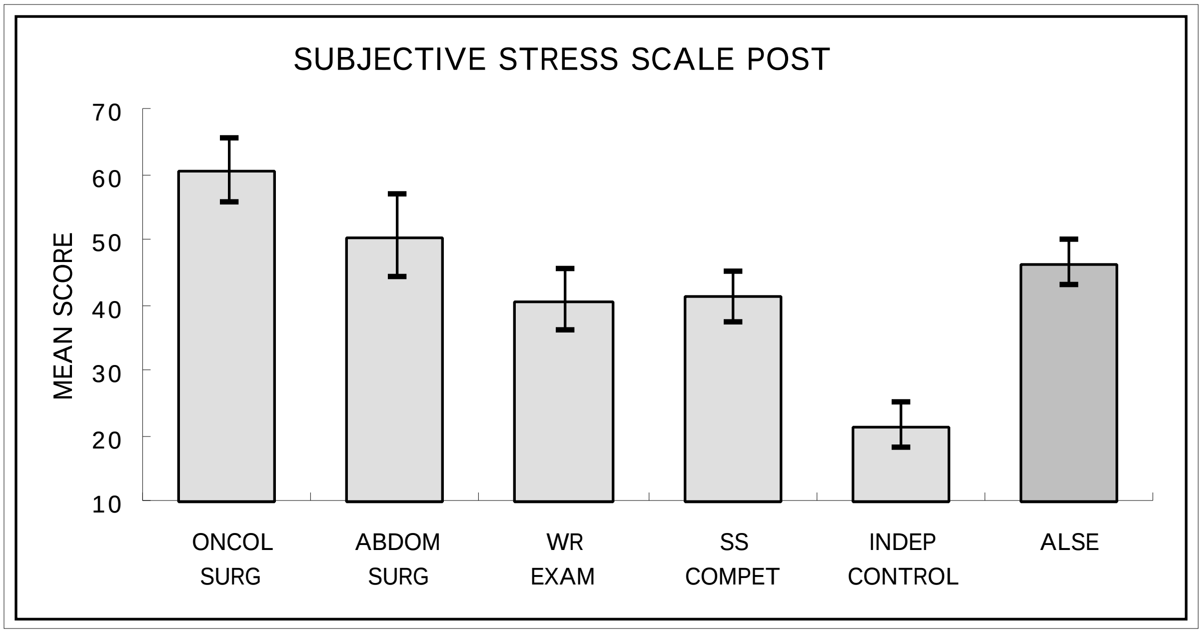


Figure 35. Comparison of mean post-stress subjective stress scores.

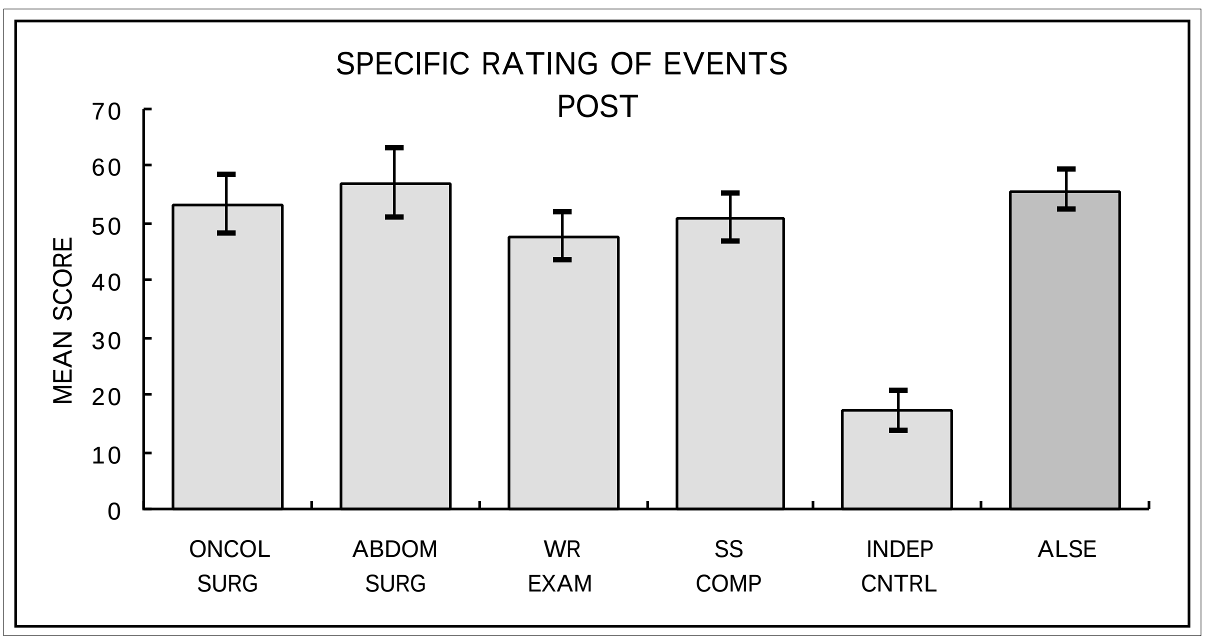


Figure 36. Comparison of mean post-stress specific rating of events scores.

A mean pre-stress specific rating of events scores histogram for the five referent protocols is not available.

The amylase assay data were compared with data from two referent protocols (Blewett, Redmond, Fatkin, Popp, & Rice, 1995; Hudgens, Chatterton, Torre, Fatkin, & King, 1990). Each of these referent protocols included a pre-stress measurement and a post-stress measurement. The two referent protocols for the present evaluation are

1. L-P DECON—Litter Patient Decontamination Group—soldiers in MOPP4 processed 40 litter-borne, fully clothed mannequins (155-pound average weight) from a triage area to a hotline.

2. PARA JUMP—Parachute Jump Group—a group of inexperienced males performed their first parachute jump. Amylase levels were measured immediately upon touchdown.

The L-P DECON and PARA JUMP protocols represent low to moderate pre-stress responses and moderate to high stress levels, post trial (see Figures 37 and 38).

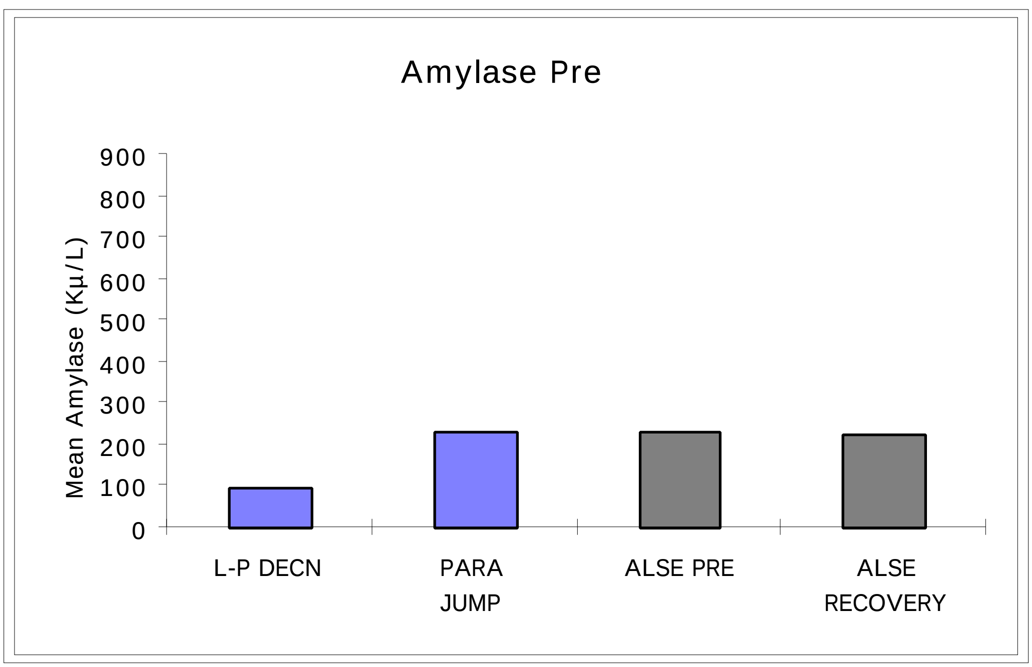


Figure 37. Comparison of mean pre-stress amylase scores.

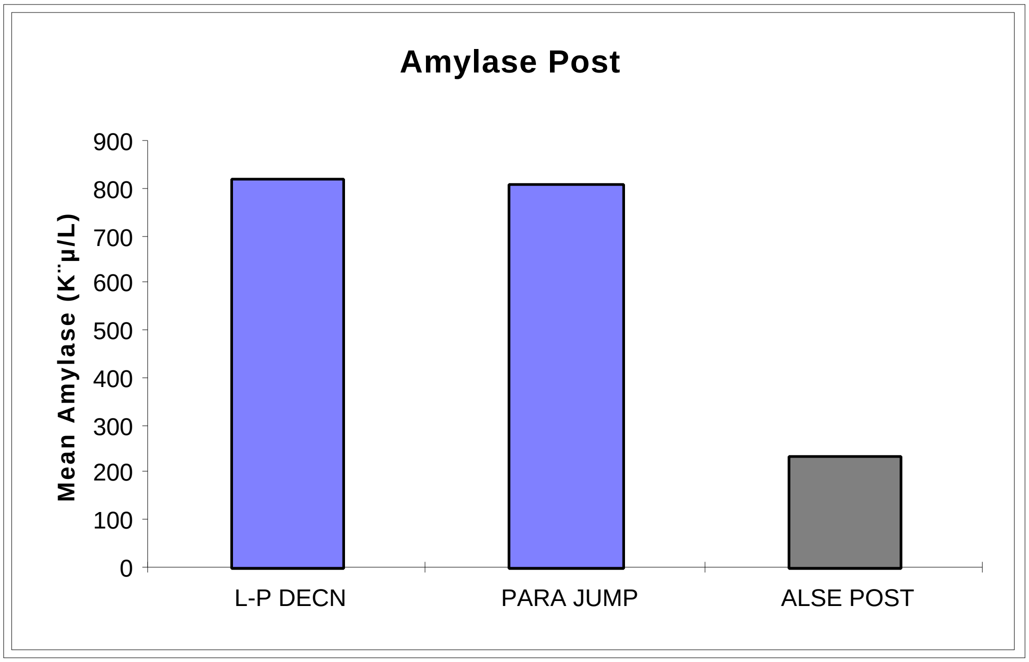


Figure 38. Comparison of mean post-stress amylase scores.

Post-Mission Interviews

Each crew was interviewed separately within 15 minutes of leaving the AH-64CMS cockpits. The format of the interviews was developed by and was shared courtesy of the Army Research Institute for the Behavioral Sciences (ARI) Unit at Ft. Rucker, Alabama, and Dr. Robert Wright of that unit (Wright, Hartson, & Couch, 1996). The investigators conducting the interviews focused on eliciting problems that each individual aviator encountered and what, if any, components or combinations of equipment worn were associated with or caused those problems. This was a somewhat formal opportunity for the subjects to “blow off a little steam” regarding their discomforts during the trials, interferences the ensemble may have caused between themselves and the cockpit, and what improvements might make the job a little easier for them. The subjects’ responses and commentary were transcribed in spreadsheet form. Equipment citations as well as discomfort citations were counted. Figure 39 summarizes those citations. Not shown in the figure is the fact that the subjects used the general term “bulk” or words to that effect 137 times, when citing individual components, combinations of components, or the entire ensemble.

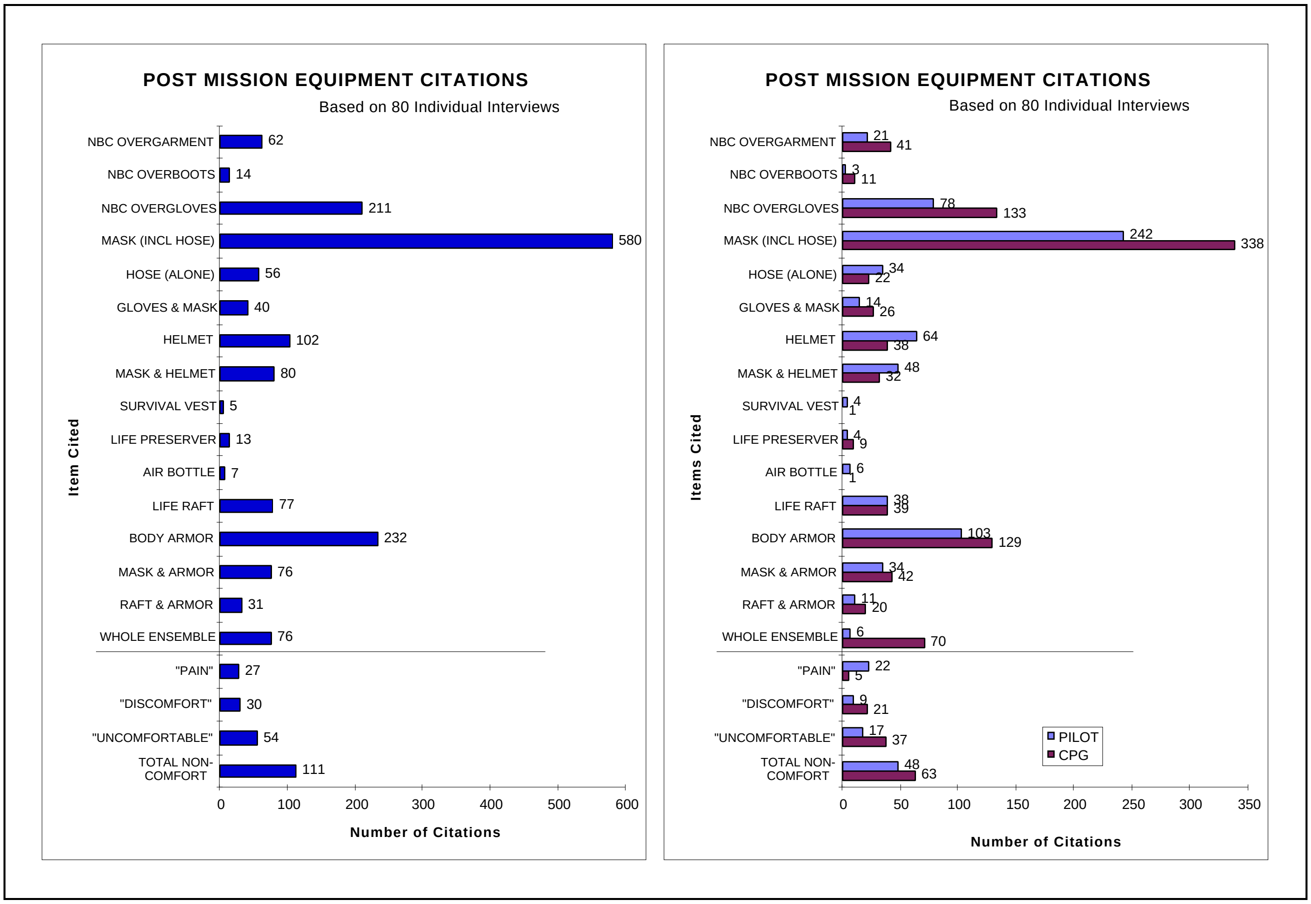


Figure 39. Post-mission citations—equipment and non-comfort.

The ensemble items and combinations of items citations were subjected to a hierarchical log linear (frequency) analysis. Trial type including familiarization trials and the items were the factors analyzed. Statistically significant one-way effects for trial type and items were found, but no two-way association was noted for trial type by ALSE item ($p < .05$). Table 29 shows the number of citations by trial type.

The distribution of items appears to be approximately even across the trials except, of course, the VR trial in which no over-water equipment was worn. The VR trial may be responsible for a significant effect of trial type. The mask was overwhelmingly cited in all trials and is probably responsible for the finding of significance; however, the overgloves and body armor also have a high number of citations.

The ARI interview questionnaire format consists of the following sections:

- Physical
- Visual
- Tactile
- Comfort
- Speech
- Mission

Table 29

Post-Mission Citation of ALSE Items by Trial Type

Item	Familiarization	FirstR	SecondR	VR
Overgarment	14	15	17	16
Overboots	5	3	1	5
Overgloves	76	54	46	35
Mask	167	155	134	124
Gloves & mask	19	9	7	5
Helmet	24	29	22	27
Mask & helmet	18	24	16	22
Survival vest	2	1	1	1
Life preserver*	4	3	4	0
HEEDS bottle*	2	2	2	0
Life raft*	26	16	22	0
Body armor	68	63	47	54
Mask & armor	17	23	17	19
Raft & armor*	10	5	10	0

* Over-water items not worn during VR trials

Figure 40 summarizes the number of citations made by the subjects in the various sections.

Note that an overwhelming number of citations were lodged against the mask, with the helmet-mask combination being cited heavily in the section concerning comfort. Note also that the section on speech, primarily intelligibility, hearing and being heard, was apparently not a problem. Further, the armor plate was cited heavily in the physical section.

The mission section of the interview questionnaire inquires about judgments of the subject's own performance and is subdivided into

- <u>Situational awareness</u>
- Radio procedures
- Navigation
- Reporting
- <u>Crew coordination</u>
- Target acquisition
- Target identification
- Target Engagement
- Weapons operation
- <u>Flight tasks</u>
- Optical relay tube, heads out
- Optical relay tube, heads down
- <u>Additional time for tasks</u>
- Alternate heads-up display techniques

The items that are underlined are deemed the more critical items in terms of overall mission performance. For the underlined items, Figure 41 shows the separate citations from the pilots and CPGs.

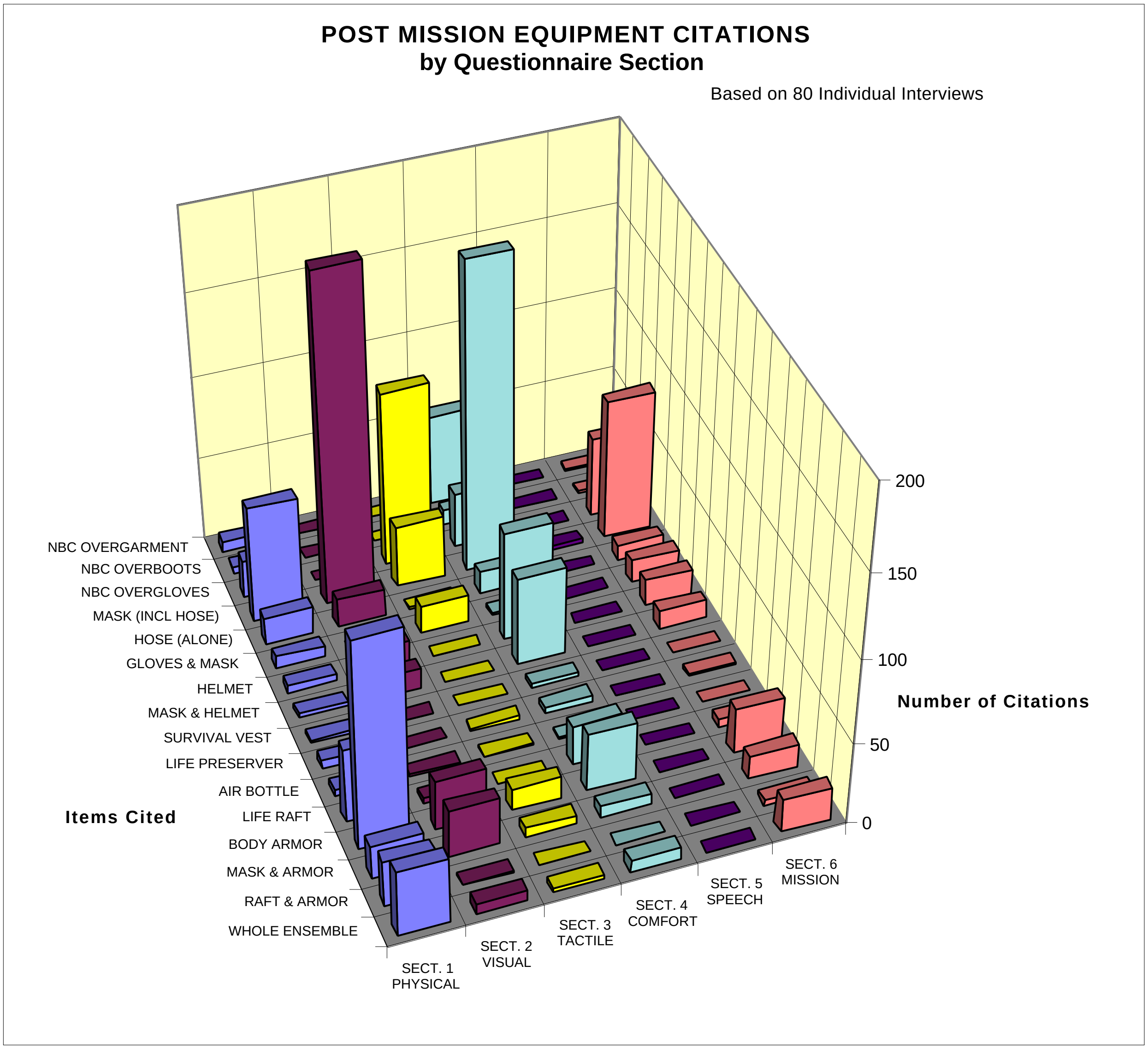


Figure 40. Post-mission citations by ARI questionnaire section.

Again, the mask, the mask and helmet in combination, and the armor plate as well as discomfort appear to be foremost in the opinions of the subjects. The CPGs had more citations than did pilots.

The counts of the major components and combinations of components were subjected to a hierarchical log linear (frequency) analysis. Trial type was the factor analyzed. Statistically significant one-way effects for trial type but no two-way associations ($p < .05$) were found. A trial type effect is to be expected for the VR trial and the citations for the over-water components, as they were not worn for the VR trials. The rest of the components and combinations were rather evenly distributed across trials.

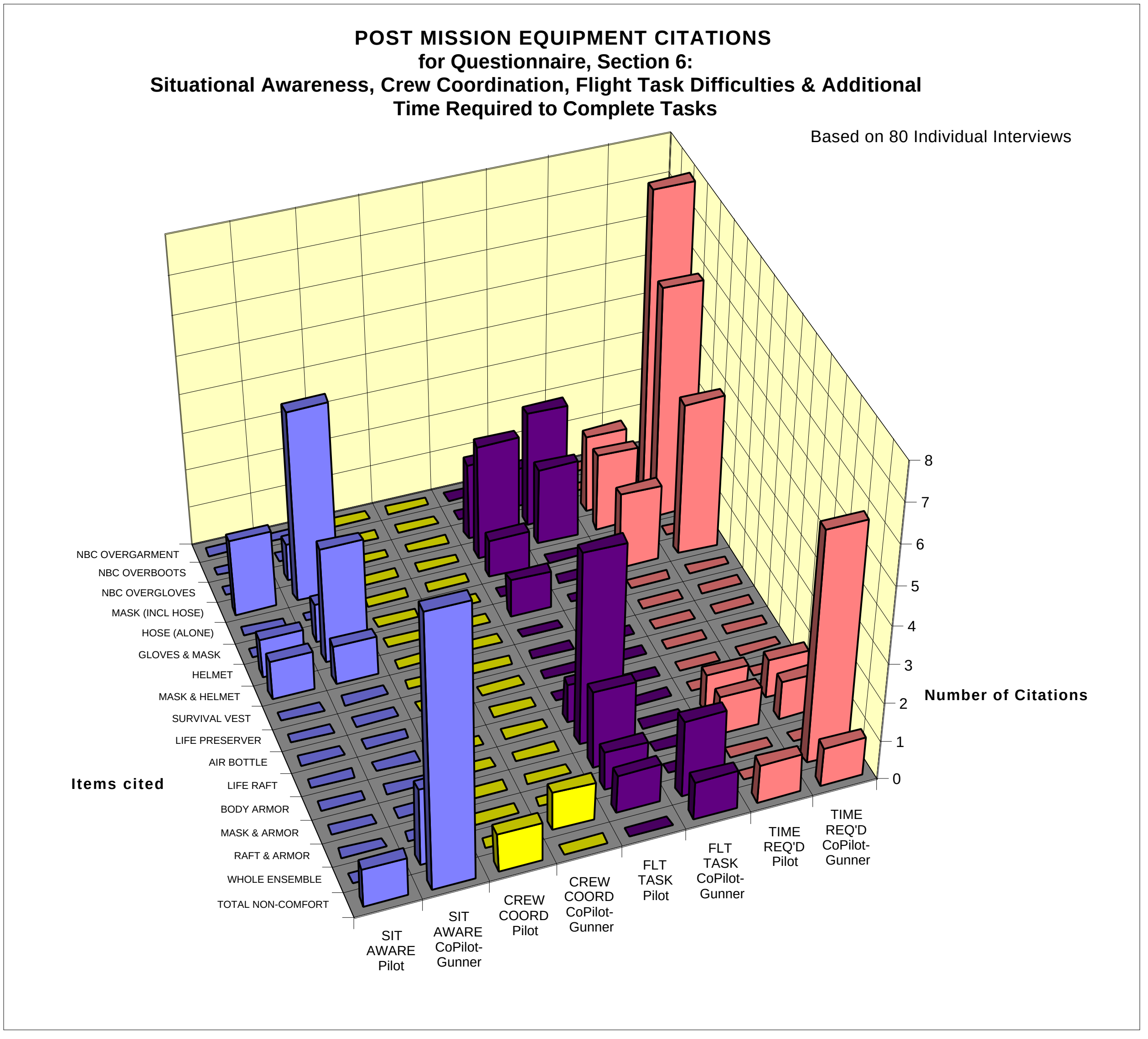


Figure 41. Post-mission citations from the ARI questionnaire mission performance section.

DISCUSSION

Behavior-Anchored Ratings

Evaluators

It would have been ideal if three or four qualified evaluators had been present for the entire experiment. Analysis of the overlapping trials between Evaluator No. 1 and 2 and between Evaluator No. 2 and 3 demonstrated that the scoring that they performed was essentially the same. As we shall see, however, the variations in the scores from crew to crew and from trial to trial were too large.

As a hedge against the collapsing or aggregating of all the evaluators' scores for the rest of the log linear analyses performed, a sampling of the 13 basic quality scores was subjected to hierarchical log linear (frequency) analyses for the scores of Evaluator No. 2 alone. Subjects, crews, trial sequence, and trial type were the factors analyzed. No statistically significant main effects for any of these factors ($p < .05$) were found.

Log Linear Frequency Analyses

None of the log linear analyses employing the behavior-anchored rating scores yielded a statistically significant main effect finding for any of the factors examined. It is believed that the distribution of these scores, regardless of how they were cross tabulated for the various factors was too large to be a sensitive measure. Figure 42 is a computation of the standard deviations of the scores, of the combined evaluators, for the 13 basic qualities. It shows standard deviations for all the subjects as well as for the pilots and CPGs alone.

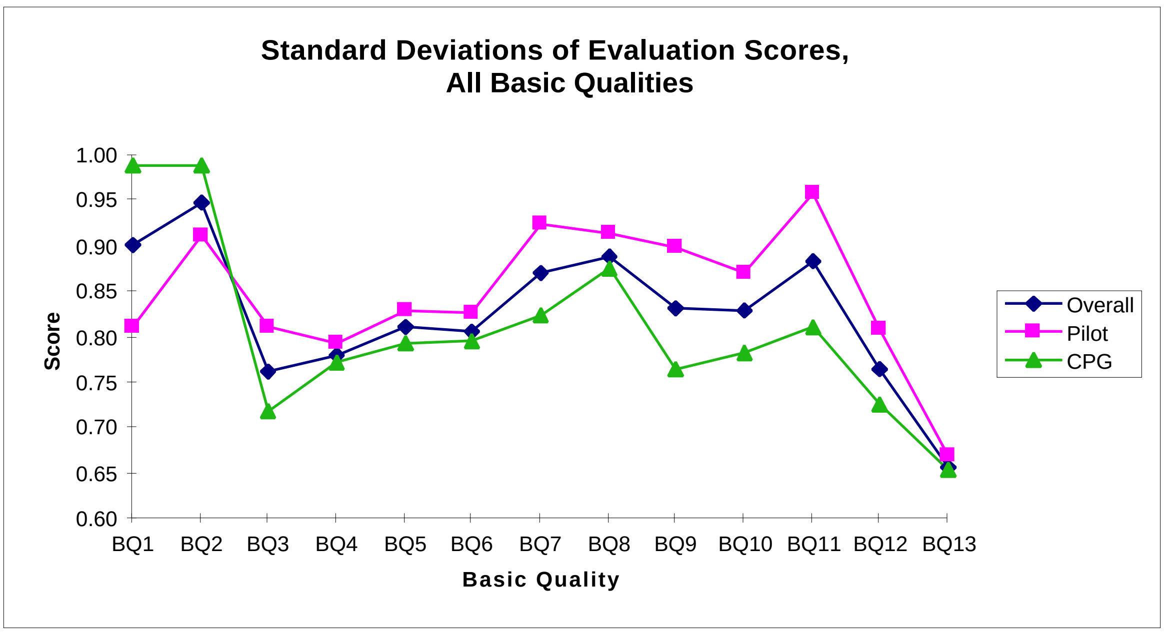


Figure 42. Standard deviations of behavior-anchored rating scores.

Figure 43 shows the standard deviations of the aggregated scores, of the combined evaluators, for the 13 basic qualities, for the third sequential trial alone. This third trial has what appears to be the lowest set of standard deviations among trial sequence or type. The situation is the same as discussed before.

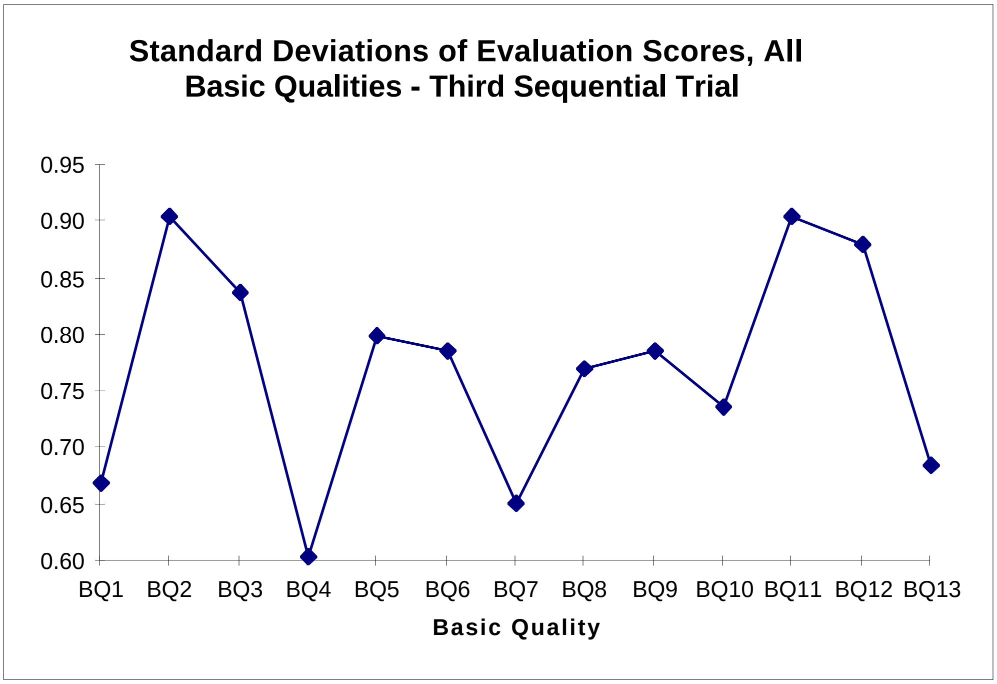


Figure 43. Standard deviations of rating scores for the third sequential trial.

Assuming any kind of normality for the scores distributions, one standard deviation, or 67% of the scores, even for the lowest value, Basic Quality 13, exceeds the bounds of one point on a Likert scale of seven points. If three standard deviations account for 98% of the distribution, then ideally, the computed standard deviation of any of the basic qualities, in order to be sensitive, ought to be around one-sixth or 0.167.

When viewed graphically, the rating scores arrayed in terms of trial sequence and trial type appear to exhibit a rather interesting set of trends, even if the findings were not statistically significant (see Figures 44 and 45).

With the aggregate of scores arranged by trial sequence, without regard for where the variation trial appeared in the order of presentation, it would appear that there was a general increase in the values, signifying better performance according to the general hypothesis behind the Likert scales. This increase is reasonable in the classical sense of more trials, more learning, or perhaps more to the point in this case, more accommodation to the ALSE ensemble. When the aggregate of scores is arranged by trial type, the apparent trend is not intuitive. Recall that in the variation trial, the subjects did not don the over-water components of the ensemble (i.e., the life raft, flotation vest, and HEEDS bottle). The life raft on the back, when not worn, left the aviators with approximately an inch more room between the seat back and the instrument panel and controls, which was greatly appreciated in the aviator's post-trial interviews. The reduced

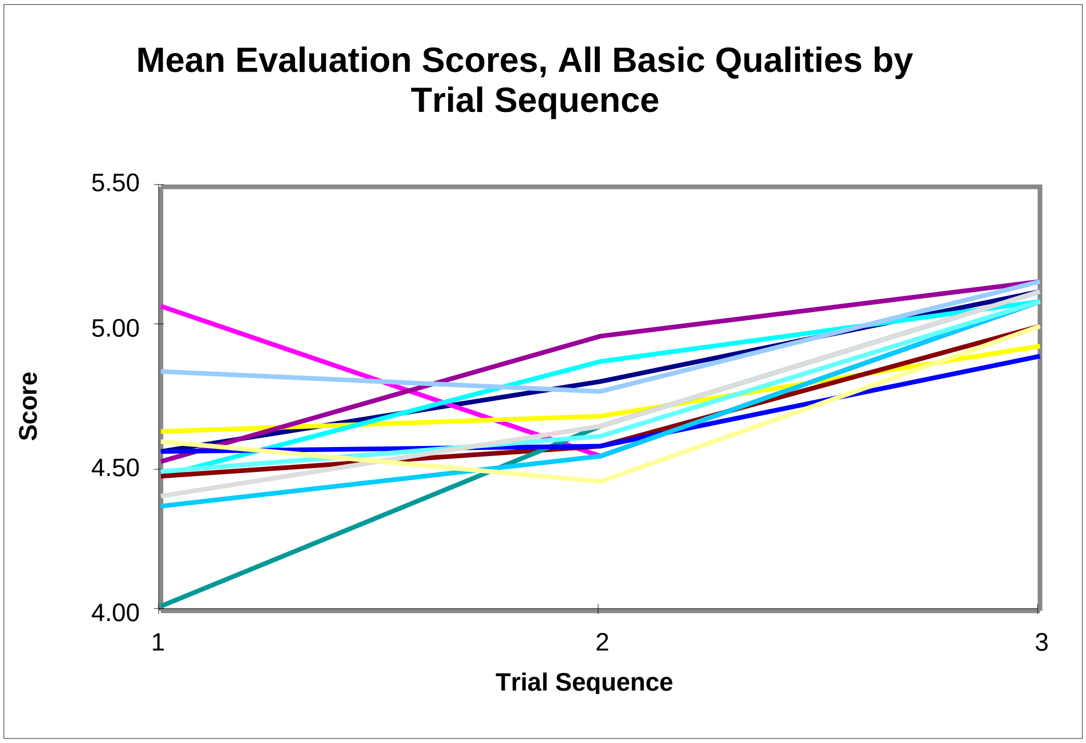


Figure 44. Scores by trial sequence.

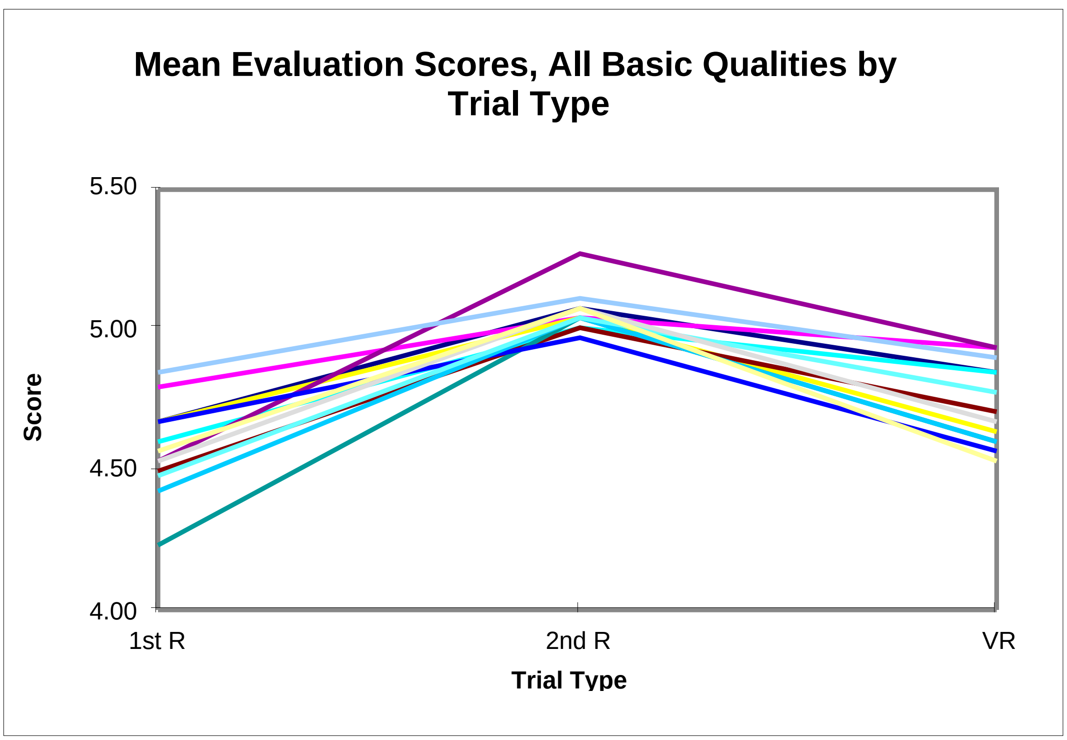


Figure 45. Scores by trial type.

ensemble should presumably have made the job easier and performance better. Yet the scores reflect that, in general, the subjects faired better in the second trial for record, and their variation trial scored somewhere in between the first and second fully encumbered trials. Surely they did not have to accommodate to the lesser encumbered, that is, easier ensemble, all over again. Had the subjects performed considerably better in the variation trial, one might consider the renowned Hawthorne Effect (Urwick & Brech, 1965) to be a factor because the investigators were doing something for the aviators—encumbering them to a lesser degree. It may be that for the trial where the subjects were dressed with fewer encumbering components, they decided that they could "take it easy this time" and thus performed more poorly than they could have. It is also felt that the evaluators, who regularly evaluate and rate other aviators, were somewhat reluctant to go the extremes of the rating scale. Of all the scores for all 13 battle qualities, only one 7 was posted, no 1s and only three 2s were posted. The subjects knew intellectually that they were scoring the ALSE and not the aviators, but it is difficult to keep that subjective element separated when observing the aviators at work in the mission. Dennis K. Leedom[7] has suggested that the use of "Critical Incident" language as related to check-ride evaluation procedures, placed in the behavioral anchors text, may serve to remove a degree of ambiguity from the scoring as well (personal communication with the first author, March 1998).

Because of the trend noted previously, an exercise removing the variation in the evaluation scores attributed to trial sequence was performed. That is, the scores were adjusted by subtracting the average difference between the first and second trials in the sequence from the second trial, and subtracting the average difference between the first and third trials in the sequence from the third trial, for the 13 basic qualities. This "flat-lined" the scores when arrayed by trial sequence. When arrayed by trial type, the scores exhibited the same characteristics as seen before but with the effects of sequence or accommodation removed (see Figure 46).

For the purposes of this exercise, a rather risky assumption is that the adjusted scores now lie along an equal interval scale (they certainly no longer lie along a Likert scale), allowing ANOVAs to be performed.

Each of the adjusted basic quality scores was subjected to a series of one-way ANOVAs with subjects, crew, seat (pilot or CPG) and evaluators each as the between-groups effects.

---

[7]Lead researcher for ARI in development of the Crew Coordination Evaluation Methods and Materials.

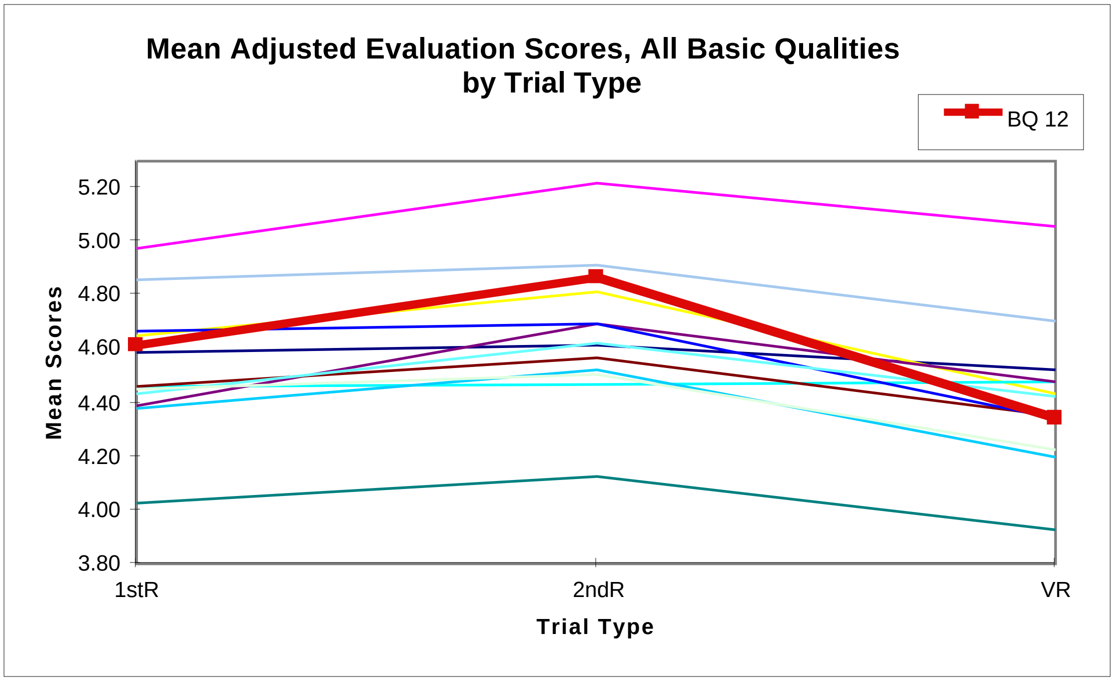


Figure 46. Mean adjusted scores by trial type.

A statistically significant effect was found for each of the basic qualities (BQ) scores except BQ 8 with respect to subjects ($p < .05$). Tukey post hoc tests were applied to the data. The results were typically 12 to 16 subjects lumped (not significantly different from one another) together, with an overlap of 6 to 12 subjects in the next non-significance group. This is to be expected when attempting to separate a large group of means.

A statistically significant effect was found for each of the basic qualities scores with respect to crews ($p < .05$). Tukey post hoc tests were applied to the data. The results were typically five to seven crews lumped (not significantly different from one another) together, with an overlap of four to six crews in the next non-significance group. This is to be expected when attempting to separate a large group of means.

No statistically significant effect was found for each of the basic qualities scores with respect to seat ($p < .05$). There were no differences in any of these adjusted scores with respect to seat.

A statistically significant effect was found for 8 of the 13 basic qualities scores with respect to evaluator ($p < .05$). Tukey post hoc tests were applied to the data. In all but one (BQ 2) of the eight significance instances, Evaluators 2 and 3 were lumped together. Evaluators 1 and 3 were never grouped together.

Each of the adjusted basic quality scores was subjected to one-way ANOVAs with trial type as the between-groups effects. A statistically significant effect was found for only one of the 13 basic qualities scores, BQ 12 (emphasized in Figure 46), with respect trial type ($p < .05$). Tukey post hoc tests were applied to the BQ 12 data (see Table 30), which point to the difference in the VR and SecondR trials as yielding a significant statistic. In other words, this is the one instance when the difference in the means was large enough and the amount of variation around these means was small enough to make a difference. Discounting this one exception, there were no differences in any of these adjusted scores with respect to trial type.

Table 30

Results of Tukey Post Hoc Test, Basic Quality 12, for Trial Type

Trial type	VR	First R	Second R
N	30	30	26
Means	4.36	4.61	4.86
Tukey			
$p < .05$			

Means underscored by the same line are not significantly different.
Means not underscored by the same line are significantly different

Figure 47 depicts the standard deviations of the adjusted scores for the 13 basic qualities, overall and by trial type. They, like the unadjusted scores, range from slightly over a half point to close to one full point (or just over in one case) on the 7-point scale. Three standard deviations would then range from one and a half to three points, too wide a range to discriminate. Figure 48 depicts this for both the unadjusted and the adjusted scores for BQ 12, and Figure 49 has the adjusted scores' standard deviations for the secondR and VR trials superimposed.

BQ 12 concerns advocacy and assertion where "...crew members advocate a course of action they consider best even when it may differ with the one being followed or proposed" (U.S. Army Aviation Center, 1992, Sect. 6). There is not very much rationale for why this particular BQ should exhibit the largest mean difference, 0.50 or half a point on the 7-point scale, for the VR and secondR trials, of all the BQs arranged by trial type, unless advocacy happened to be a particularly important point in the minds of the evaluators during their scoring.

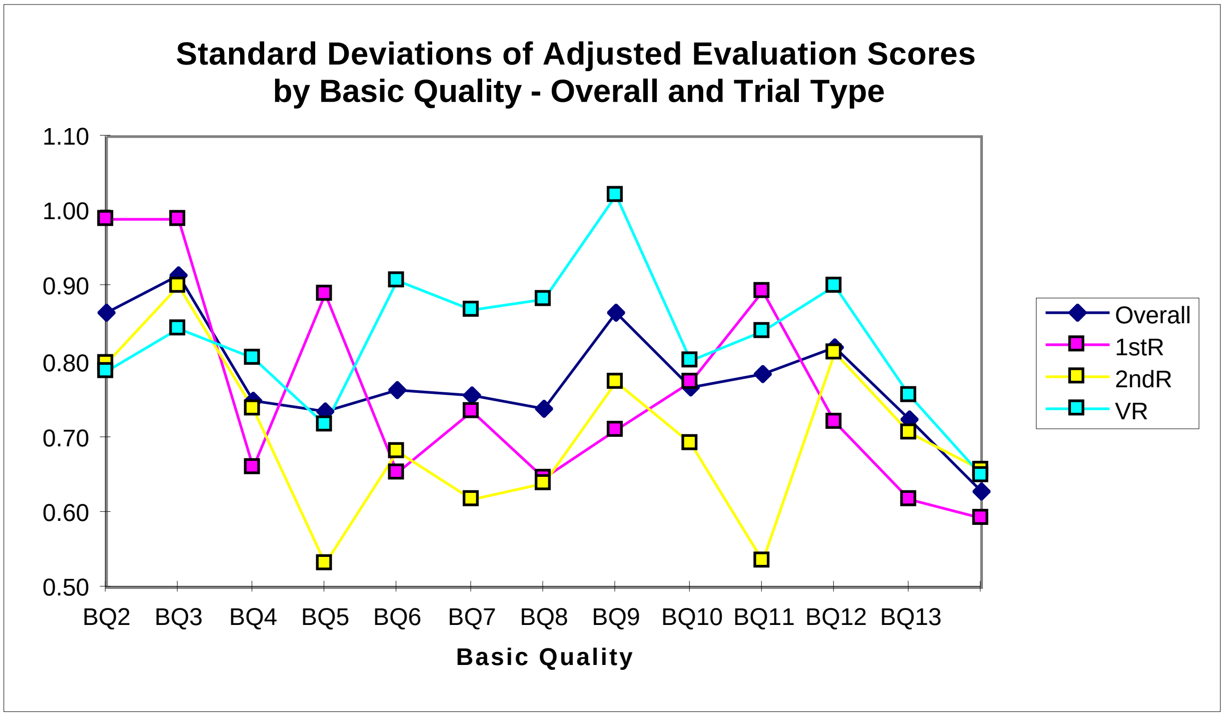
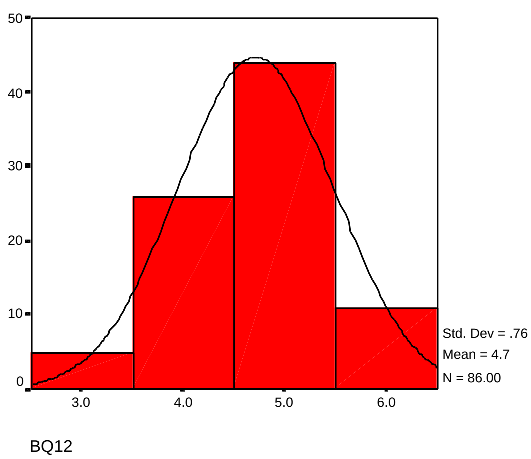


Figure 47. Standard deviations of adjusted rating scores, overall and by trial type.



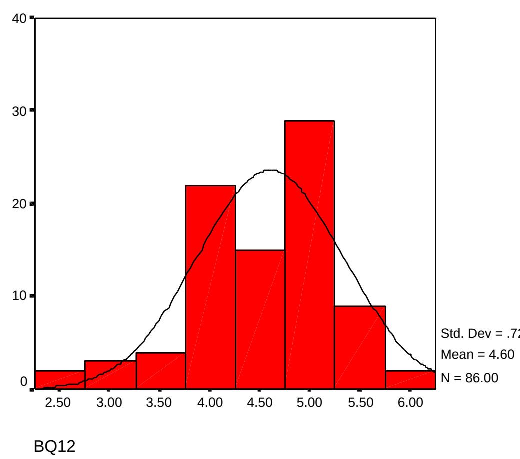


Figure 48. Histogram and normal curve for unadjusted and adjusted scores—Basic Quality 12.

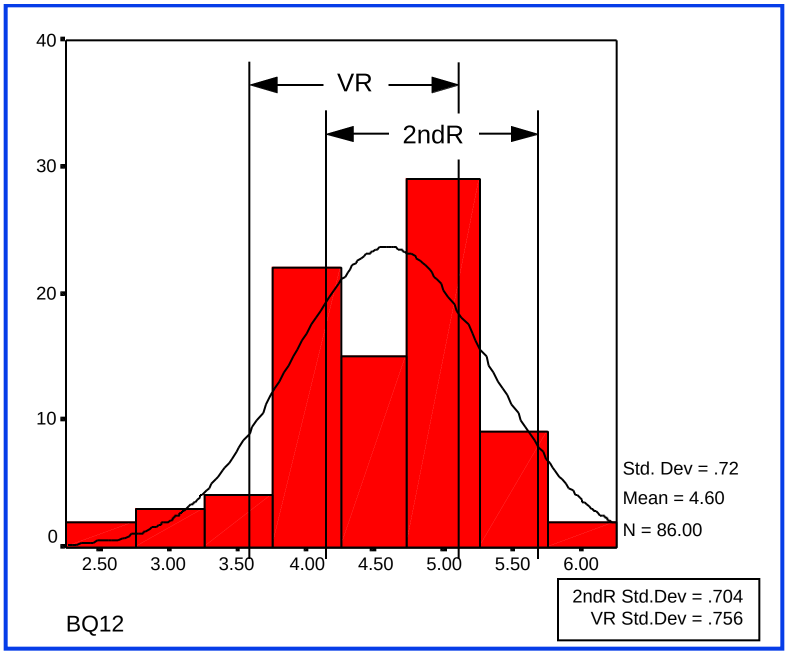


Figure 49. Histogram and normal curve for adjusted scores with secondR and VR trial standard deviations—Basic Quality 12.

The trend of the having the SecondR trial, even with the adjusted scores as the criteria, be the best of the three trials for record instead of the VR trial is readily apparent. An explanation for this is not apparent. In the post-trial interviews, all the subjects, to a man, expressed that the VR trial was a relief from being fully encumbered in the preceding trial(s).

Aviator Experience

The group of subjects participating had a rather complete range of experience in both total flying hours logged and in Apache flying hours logged. It might be expected that in general, the subjects would perform in proportion to their level of experience, that is, they would have higher scores if they had more experience and lower scores if they had a lesser number of hours in their logbooks. Perhaps a regression could be applied. The log linear analyses examining experience categories, both total and for the AH-64, weighted by each of the 13 basic quality scores, failed to conclude that a relationship exists. A sampling of basic quality scores against raw flying experience, not categorized, also failed to find a relationship. The aggregated scores were subjected to a Spearman ρ correlation computation against both total and AH-64 experience, categorized and raw hours (see Table 31).

Table 31

Spearman's ρ Correlations: Basic Quality Scores by Flying Experience

	BQ1	BQ2	BQ3	BQ4	BQ5	BQ6	BQ7	BQ8	BQ9	BQ10	BQ11	BQ12	BQ13
Total hrs.cat	.093	.139	.018	.075	.126	.059	.083	.007	.106	.121	.095	.056	-.144
Total hours	.074	.106	.001	.038	.106	.061	.067	.006	.099	.124	.098	.027	-.180
AH-64 hrs.cat	.083	.154	.093	.065	.161	.117	.143	.037	.155	.171	.197	.094	-.131
AH-64 hours	.088	.195	.102	.067	.178	.138	.170	.059	.155	.199	.222*	.126	-.134

* Correlation is significant at the .05 level (two-tailed).

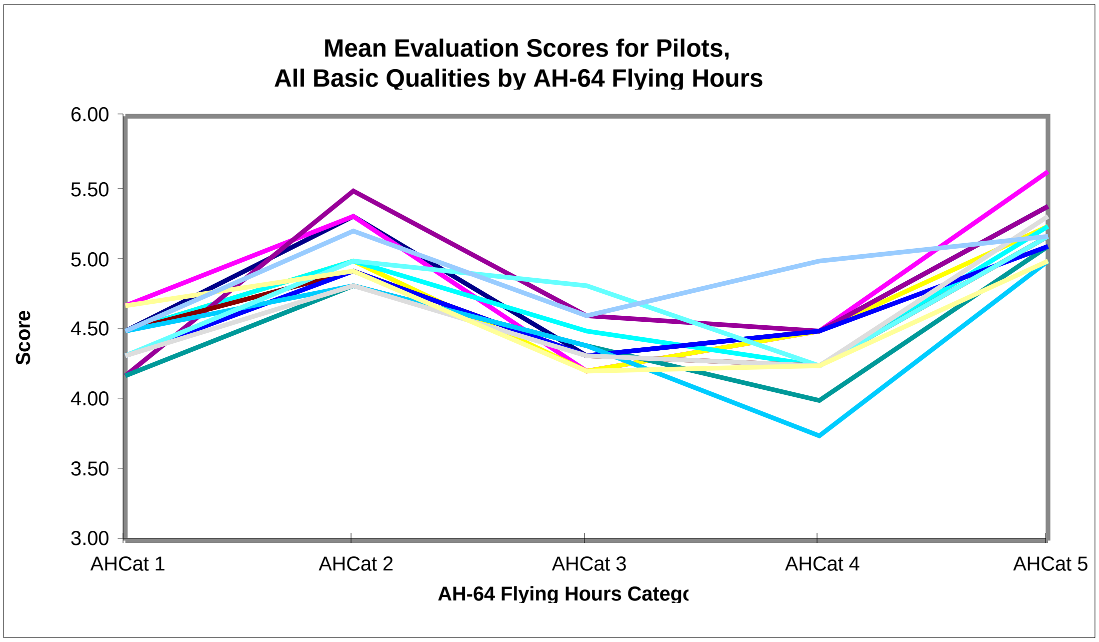
Flying experience does not at all correlate with the behavior-anchored scores. While not significant, it is curious that BQ 13, "Crew-level after-action reviews accomplished," had all negative coefficients. Figure 50 graphically illustrates why no correlation is present.

Post-Mission Interviews

While there was an abundance of equipment citations and complaints, there did not appear to be a connection between the equipment components cited and any particular trials encountered. Pearson's ρ correlation coefficients were computed for the number of times each component or combination of components was cited and each of the 13 basic quality adjusted scores (aggregated) was rendered for each subject and trial for record (see Table 32). No correlation coefficients approach the customary .70 level that signifies a potentially definitive relationship among the variables.

Stress Assessment of Aviator Subjects

An evaluation of the stress experience revealed that pilots wearing the ALSE were only moderately stressed following the test sessions. The Time x Subscore interaction found in the MAACL-R combined data is readily explained by the fact that as one's anxiety, hostility, depression, and negative affect levels increase, one's positive affect level will naturally decrease. This relationship can easily be seen in Figure 12.



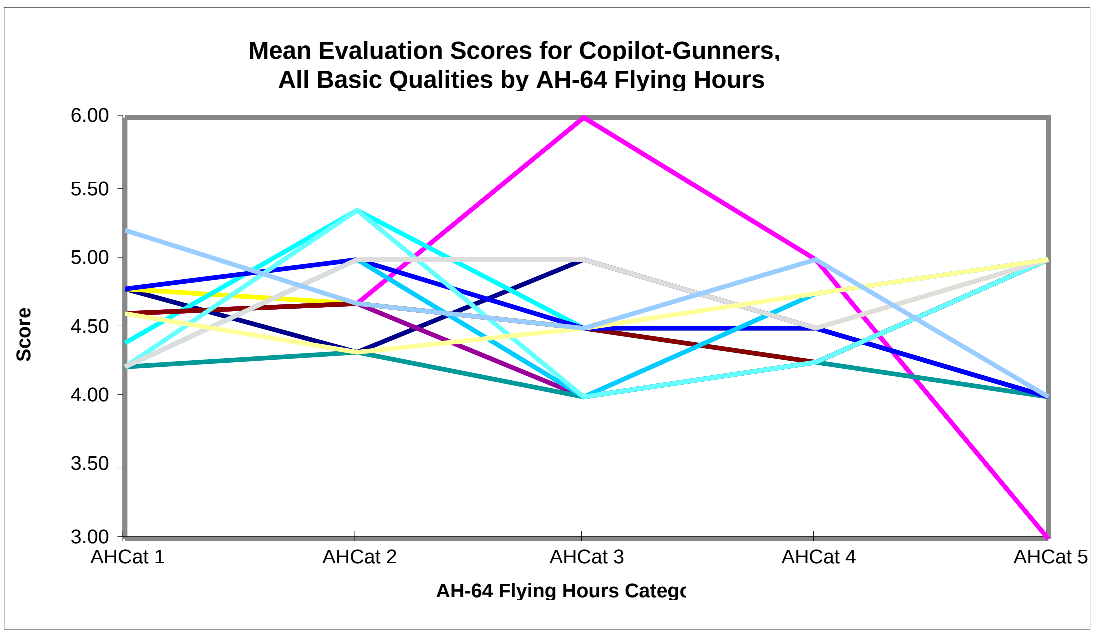


Figure 50. Behavior-anchored scores by AH-64 flying experience—pilot and CPG.

The depression subscore showed a Trial x Time interaction. This is consistent with previous research (Blewett, Redmond, Fatkin, Popp, & Rice, 1995; Mullins, Fatkin, & Patton, 1998) in that higher levels of depression on the first day of testing typically result when soldiers' expectations exceed their actual performance.

Table 32

Spearman's ρ Correlations: Adjusted Basic Quality Scores by Equipment Citations

	BQ-1	BQ-2	BQ-3	BQ-4	BQ -5	BQ-6	BQ-7	BQ-8	BQ-9	BQ-10	BQ-11	BQ-12	BQ-
Overgarment	0.013	-0.070	0.021	-0.042	-0.071	-0.023	-0.077	-0.136	-0.149	-0.233*	-0.094	0.011	0.1
Overboots	-0.221*	-0.246*	-0.189	-0.221*	-0.330*	-0.168	-0.198	-0.191	-0.066	-0.188	-0.138	-0.114	-0.1
Overgloves	0.161	0.104	0.167	0.089	0.070	0.072	0.038	0.019	-0.117	-0.136	0.021	0.040	0.1
Mask	-0.032	-0.110	-0.029	-0.150	-0.181	-0.059	-0.180	-0.109	-0.160	-0.179	-0.137	-0.039	-0.0
Gloves & mask	-0.037	0.057	-0.002	-0.104	-0.073	-0.089	-0.001	-0.126	-0.154	-0.074	-0.071	-0.026	-0.1
Helmet	0.006	-0.213*	0.022	-0.030	-0.020	0.001	-0.147	-0.069	-0.166	0.018	-0.148	-0.007	0.0
Mask & helmet	-0.096	-0.275*	-0.069	-0.080	-0.135	-0.151	-0.251*	-0.177	-0.349**	-0.160	-0.295**	-0.096	0.0
Survival vest	-0.094	0.001	-0.121	0.014	-0.015	-0.130	0.038	-0.118	-0.119	0.015	-0.114	-0.281**	-0.1
Life preserver	-0.216*	-0.112	-0.207	-0.133	-0.196	-0.260*	-0.151	-0.129	-0.285**	-0.086	-0.206	-0.358**	-0.1
Air bottle	-0.050	-0.066	-0.073	-0.064	-0.081	-0.012	-0.040	-0.001	-0.108	0.084	-0.067	-0.251*	0.0
Life raft	-0.037	-0.115	0.042	0.013	-0.042	0.012	-0.003	0.054	0.037	0.010	-0.022	0.105	0.0
Body armor	-0.075	0.072	-0.071	-0.056	-0.147	-0.187	-0.100	-0.140	-0.129	-0.316**	-0.178	0.019	-0.1
Mask & armor	-0.206	-0.072	-0.164	-0.178	-0.310**	-0.227*	-0.139	-0.224*	-0.157	-0.323**	-0.242*	-0.086	- 0.297**
Raft & armor	-0.135	-0.097	-0.049	-0.038	-0.176	-0.089	-0.046	-0.024	0.083	-0.130	-0.063	0.136	-0.2

* Correlation is significant at p<.05 (two-tailed).
** Correlation is significant at p<.01 (two-tailed).

Psychological and Physiological Stress Assessment

Both the stress perception and salivary amylase measures were used to quantify individual reactions to potential physical and mental stressors using noninvasive methods with no interference with the performance of the aviators. When administered concurrently, as in the present study, these measures provide diagnostic information that helps to identify specific components of stress. The amylase measure has provided data that indicate “how much” or “how high” the individual’s physiological stress reactions are, yet do not give information about which stress component is at work. The stress perception measures such as the MAACL-R subscales and the Subjective Stress Scale provide information about the probable source of stress.

An evaluation of the stress experience revealed that pilots wearing the ALSE were only moderately stressed following the test sessions. The Time x Subscore interaction found in the MAACL-R combined data is readily explained by the fact that as one’s anxiety, hostility, depression, and negative affect levels increase, one’s positive affect level will naturally decrease. This relationship can easily be seen in Figure 12.

The depression subscore showed a Trial x Time interaction. This is consistent with previous research (Blewett, Redmond, Fatkin, Popp, & Rice, 1995; Mullins, Fatkin, & Patton, 1998) in that higher levels of depression on the first day of testing typically result when soldiers’ expectations exceed their actual performance.

Comparison With Other Protocols

The data obtained in the current study were compared with data obtained in previous studies conducted by ARL. These comparisons provide a method for estimating the relative stress experienced in a given situation. The two surgical protocols represent a relatively high stress level; the written exam and the soldier competition protocols represent a relatively moderate level of stress, and the independent control protocol represents a condition ranging from no stress to low stress.

The stress perception measures completed before testing (pre) and those obtained 1 hour after testing were equivalent to the INDEP CNTRL group, indicating that the pilots were experiencing little or no stress at those time points. However, post measures more closely resemble the levels of stress experienced by the WR EXAM and SS COMP groups. Both of these groups indicate low to moderate levels of stress. Moreover, the stress induced by the WR EXAM and the SS COMP groups is similar in nature to that of the pilots in the current study in that it creates a competitive situation in which the individuals are in control of the outcome.

Although the aviators' salivary amylase concentrations were significantly higher during the trials than during their pre-trial or recovery time points, this increase does not indicate that the aviators were physically stressed during their trials. When the data are compared with other studies identified as moderate to high stress scenarios, L-P DECON and PARA JUMP protocols, the actual values of amylase concentration for the aviators are acknowledged as relatively low. A recent study was conducted using both experienced flight instructors and relatively inexperienced cadets performing real and simulated British Aerospace Hawk MK 51 flights (Ylonen, Lyytinen, Leino, Leppaluoto, & Kuronen, 1997). The effects of psychological workload on physiological symptoms, such as heart rate, were investigated. No statistical differences were found between the groups, and the only differences in heart rate occurred during precision instrument final approaches to landing, in minimum weather, which was considered a cognitively loaded portion of the flight. As in the current study, the changes in physiological responses were not attributed to a high level of physical workload.

Further examination of the stress levels obtained from the psychological measures reveals that the moderate stress experienced by the aviators was primarily attributable to a sense of frustration or discouragement regarding their performance during adverse conditions. The aviators were required to perform while wearing an uncomfortable ensemble in a confined crew station, within a battlefield mission and related threat environment. Hence, the stress experienced was attributable to the psychological or emotional component. Although a moderate level of psychological stress is typical during a state of arousal or vigilance, it was not dramatic enough to create an influence on aviator performance. The low levels of anxiety reported by the aviators indicate that they were certain about the tasks they were asked to perform but may have been somewhat frustrated or critical of their own performance.

Stress Assessment and Evaluation Scores

Spearman's ρ correlations were computed for the post-psychological and post-amylase measures and the 13 basic qualities scores. The computed coefficients did not have any patterns of association with any particular measure, and no coefficients reached the .70 level, signifying a potentially definitive relationship (see Table 33).

Measures of Effectiveness and Flight Data

This set of measures, when analyzed, failed to discriminate among either trial sequence or trial type. It does represent a baseline of sorts that may be useful in the assessment of prototype or future ensembles and components. One set of measures stands out in an anecdotal fashion:

Table 33

Spearman's ρ Correlations: Basic Quality Scores by Stress Assessment Measures

	BQ-1	BQ-2	BQ-3	BQ-4	BQ -5	BQ-6	BQ-7	BQ-8	BQ-9	BQ-10	BQ-11	BQ-12	BQ-1
Subjective stress	-0.090	-0.184	-0.051	-0.034	-0.080	-0.275*	-0.128	-0.197	-0.356**	-0.352**	-0.326**	0.023	0.218
SRE	0.172	-0.194	0.166	0.131	0.067	0.038	-0.099	-0.026	-0.140	-0.022	-0.042	0.197	0.345
Anxiety	0.038	-0.046	0.124	0.043	0.141	-0.044	-0.115	0.054	-0.049	-0.018	-0.015	0.177	0.152
Hostility	-0.154	-0.197	-0.233	-0.284*	0.005	-0.114	-0.104	-0.065	-0.150	-0.101	-0.277*	-0.079	0.160
Depression	-0.022	-0.112	-0.020	-0.039	0.104	-0.136	-0.033	-0.056	-0.300*	-0.289*	-0.293*	0.022	0.284
Positive affect	0.117	0.335**	0.068	-0.028	-0.047	0.072	0.061	0.030	0.072	0.106	0.189	-0.056	-0.02
Negative affect	-0.050	-0.166	-0.085	-0.092	0.123	-0.079	-0.085	-0.040	-0.169	-0.100	-0.228	0.009	0.187
Amylase	-0.052	-0.118	-0.031	0.055	-0.106	-0.084	-0.198	-0.079	-0.185	-0.315*	-0.192	-0.124	0.208

* Correlation is significant at p<.05 (two-tailed).
** Correlation is significant at p<.01 (two-tailed).

during the initial survey of ATM tasks, before the conducted trials at WAATS, several of the surveyed Apache aviators at Ft. Hood, Texas, as well as at Ft. Rucker, Alabama, volunteered that the usual time taken to “boot up” the AH-64 was on the order of 20 minutes even with raised cockpit temperatures when the helicopters were standing on a ramp in bright sunlight. The time taken in the AH-64CMS with temperatures around 68° F, was in the neighborhood of 30 minutes.

In general, the AH-64CMS computed measures, when analyzed, were found to have significant test statistics for trial sequence, namely, the first and third trials were different, except for the events, hits, misses, and no-shots data which showed no significant differences in trial sequence nor in trial type. The only significant test statistic for trial type occurred when the means for Maximum Probability of Hit were examined; then the firstR trial differed from the VR trial. Counter to expectation, the crews (only five) whose missions were captured by the CMS all had zero hit probabilities assigned to their first fully encumbered trials for record.

Spearman’s ρ correlation coefficients were computed for the weapons engagement data and for each of the 13 basic quality adjusted scores (aggregated) rendered for subjects and their trials for record (see Table 34). Among the highest computed were the negative coefficients associated with missile shots that missed the intended targets; however, no correlation coefficients reached the .70 level, signifying a potentially definitive relationship among the variables.

Spearman’s ρ correlation coefficients were computed for the threat measures data and each of the 13 basic quality adjusted scores (aggregated) rendered for subjects and their trials for record (see Table 35). Among the highest computed were the negative coefficients associated with total and average exposure times along with the mean exposure zone per event; however, no correlation coefficients reached the .70 level, signifying a potentially definitive relationship among the variables.

In general, the data extracted from the CMS mainframe contained gaps and were incomplete across trials, often beyond the control of the instructor-operators. The job of collating these data by hand was also long and tedious. Ideally, investigators should have had the ability to extract and collate these data electronically via software programming. If further work is anticipated in this venue, it would be worthwhile pursuing this course of action.

Both the engagement measures and threat measures able to be gathered in this investigation are indicators of the crew’s performance during their missions. Additionally, it became quite apparent to the investigators, the evaluators, and even to the simulator instructor-operators that the ALSE ensembles and their attendant distractions were probably responsible for a reduction in the

Table 34
Spearman's ρ Correlations: Adjusted Basic Quality Scores by Weapons Engagement Measures

	BQ-1	BQ-2	BQ-3	BQ-4	BQ -5	BQ-6	BQ-7	BQ-8	BQ-9	BQ-10	BQ-11	BQ-12	BQ-1
Engagements	-0.091	-0.043	-0.268*	-0.180	-0.231	-0.159	-0.079	-0.156	0.072	-0.011	0.066	-0.047	-0.10
Engagements-missile	-0.157	-0.220	-0.223	-0.140	-0.429**	-0.338**	-0.188	-0.266*	-0.191	-0.142	-0.150	-0.232	-0.16
Engagements-rocket	-0.242	-0.010	-0.172	-0.103	0.036	-0.119	-0.016	-0.079	-0.069	-0.202	-0.295*	-0.090	-0.22
Engagements-gun	-0.024	-0.161	-0.193	-0.119	-0.192	-0.054	-0.097	-0.078	0.082	0.072	0.156	0.005	-0.09
Kills	0.111	-0.071	-0.071	0.093	-0.180	-0.106	0.058	-0.112	0.014	0.040	0.125	0.035	-0.05
Hits	0.035	0.168	-0.123	0.003	0.019	-0.071	0.116	-0.048	-0.101	-0.086	-0.069	-0.063	-0.04
Misses	-0.228	-0.145	-0.253	-0.290*	-0.199	-0.141	-0.198	-0.179	0.040	-0.019	-0.028	-0.124	-0.16
Percent kills	0.239	-0.037	0.155	0.286*	-0.044	0.025	0.144	0.047	0.054	0.100	0.195	0.171	0.08
Percent hits	0.062	0.201	-0.086	0.045	0.043	-0.067	0.133	-0.011	-0.054	-0.058	-0.038	-0.023	0.07
Percent misses	-0.237	-0.145	-0.124	-0.272*	-0.133	-0.065	-0.288*	-0.094	-0.044	-0.052	-0.130	-0.166	-0.08
Missile kills	0.151	0.096	0.069	0.118	-0.100	-0.017	0.180	0.015	0.090	0.069	0.166	0.067	-0.00
Missile hits	-0.052	-0.040	0.008	-0.032	-0.056	-0.086	-0.153	0.000	-0.234	0.114	-0.132	-0.214	0.04
Missile misses	-0.404**	-0.404**	-0.426**	-0.364**	-0.639**	-0.548**	-0.525**	-0.485**	-0.372**	-0.362**	-0.376**	-0.410**	-0.30
Rocket kills	-0.200	-0.222	-0.251	-0.134	-0.087	-0.193	-0.165	-0.236	-0.285*	-0.182	-0.331*	-0.274*	-0.16
Rocket hits	0.044	0.182	-0.102	-0.025	0.040	-0.038	0.162	-0.052	-0.076	-0.123	-0.040	-0.033	-0.04
Rocket misses	-0.261	-0.091	-0.113	-0.102	-0.029	-0.093	-0.103	-0.035	-0.029	-0.111	-0.277	-0.085	-0.18
Gun kills	0.118	-0.050	-0.073	0.056	-0.140	-0.037	0.023	-0.054	0.119	0.118	0.242	0.129	-0.00
Gun hits	0.059	0.009	-0.159	0.126	-0.159	-0.124	-0.041	-0.016	0.055	-0.071	0.006	0.029	-0.05
Gun misses	0.008	-0.114	-0.090	-0.105	-0.094	0.038	-0.002	0.007	0.107	0.136	0.195	0.030	-0.03

* Correlation is significant at $p<.05$ (two-tailed).
** Correlation is significant at $p<.01$ (two-tailed).

Table 35
Spearman's ρ Correlations: Adjusted Basic Quality Scores by Threat Measures

	BQ-1	BQ-2	BQ-3	BQ-4	BQ -5	BQ-6	BQ-7	BQ-8	BQ-9	BQ-10	BQ-11	BQ-12	BQ-1
Number of events	-.233	-.300	-.209	-.068	-.100	-.135	-.051	-.173	.282	.139	.019	.196	.002
Number of hits	-.102	-.116	.011	-.022	-.165	-.057	-.277*	.080	-.051	.026	-.102	-.186	-.111
Number of misses	-.136	-.234	-.077	-.091	-.296*	-.143	-.341**	-.141	.108	.127	.024	-.108	-.071
Number of no shots	-.094	-.136	-.127	.057	.087	-.006	.123	-.061	.280*	.164	.066	.238	.031
Percent shot at	-.044	-.070	-.022	-.069	-.280*	-.122	-.339**	-.082	.011	.074	.034	-.180	-.087
Total exposure time	-.374**	-.567**	-.342**	-.261*	-.309*	-.276*	-.226	-.323*	-.036	-.091	-.241	-.043	.080
Average exposure time	-.428**	-.587**	-.402**	-.374**	-.385**	-.347**	-.274*	-.390**	-.235	-.309*	-.385**	-.157	.037
Mean exposure zone	.243	.195	.282*	.187	.189	.220	.024	.091	.195	.343**	.289*	.033	.095
Sum of exposure zones	-.182	-.218	-.131	-.035	-.007	-.070	.011	-.135	.328*	.214	.094	.237	.019
Max probability of acquisition	-.009	.003	.033	-.049	.150	.048	.167	-.093	.033	.004	.052	.075	-.023
Max probability of hit	-.194	-.404**	-.107	-.127	-.095	-.136	-.146	-.252	-.026	.040	-.073	-.016	.021

* Correlation is significant at $p<.05$ (two-tailed).
** Correlation is significant at $p<.01$ (two-tailed).

collective as well as the individual aviator's situation awareness. Evidence ranged from instances of too great an exposure and time of exposure to a threat, to munitions fired into the intervening terrain, to having switches and selectors in the wrong position, to the inability to see and react to an illuminated master caution light. The National Research Council (1997) states that in aircraft simulation there is evidence of correlation between mission performance and situation awareness, but because of the many factors influencing this link, performance measures can only provide an indirect indication situation awareness. It is universally acknowledged that there is a definite relationship between performance, workload, and situation awareness albeit a complex one, and that techniques are advocated to address the quantification of situation awareness (Endsley, 1990). Any further research along these lines ought to include the influence that ALSE technology, or for that matter, the influence of any proposed new aircraft systems would have on situation awareness and especially systems whose stated purpose is to enhance aviator's situation awareness. The raising of an aviator's situation awareness in one particular information set could precipitate a reduction in situation awareness along one or more other, perhaps critical, channels.

Discussion Summary

The behavior-anchored rating scores were not found to have main effect significances. They did not discriminate among individual aviators, crews, or seat position occupied, which may be the fault of the wider than required variations in their distributions or may be that there are no real differences among individual aviators, crews, or seat position occupied. The behavior-anchored rating scores were not found to be significant in terms of trial sequence or trial type. In general, aviator performance could be judged between "adequate" (scored as 4) and "good" (scored as 5). Graphically, however, it is quite apparent that an upward trend existed in terms of the sequence of trials. This could be attributable to the subject's continued accommodation (not necessarily learning) over the three trials for record. Also very apparent was the fact that the variation trials (no over-water equipment) did not produce the highest of the scores by type of trial, when intuitively, they should have. The subjects may have been exhibiting a degree of complaisance during that trial or may have had to re-accommodate to the changed ensemble.

Intuition also tells us that the more experienced the aviator, the higher that aviator as a subject would be likely to score in the trials, and if true, the evaluation scores could be weighted and thence be more apt to discriminate. Such was not the case. There were no correlations of experience by basic quality scores higher than .222, and the distribution appears to be almost random. This is not an avenue to be pursued.

The adjustment of the evaluation scores to level the assumed effect of accommodation appeared to improve the analysis but did not quite surpass the probability boundary into significance except for BQ 12. This adjustment probably should be part of the analysis routine in any further research along this line.

The three components of the ALSE ensemble giving the subjects the most problems were the protective mask, body armor, and protective overgloves, followed by the helmet, life raft, and the combinations of helmet-mask and mask-armor. The combination of the armor in front and the raft in back caused interference between the CPG and the front seat cyclic stick, preventing full aft displacement for both crew members. “Pitch-up” maneuvers such as hovering quick stops and landing flares had to be well anticipated because of this dangerous limitation. This has been previously observed in the Apache cockpit of the ARI simulator at Ft. Rucker (Wright, Hartson, & Couch, 1996).

While the stress perception measures, salivary amylase concentrations, and psychological measures indicated a higher level of stress in the aviators immediately following trials, than did either pre-trial or recovery levels, these levels were universally lower than those recorded in other studies characterized as moderate to high stress scenarios.

The capturing of data arising from the AH-64CMS mainframe is full of gaps and not user friendly in the gathering and reduction of data. If further research or comparative assessments are to be conducted on this type of facility, automated capturing of electronic data streams of performance-related parameters should be pursued and accomplished. The data gathered tend to form a baseline to accompany the other findings of the experiment.

While not formally addressed in the investigation, evidence of situation awareness degradation, as an effect of the ensembles worn, was observed. Inclusion of this factor in subsequent research would be well advised.

CONCLUSIONS

The main-line concept of using the framework of crew coordination evaluation, behavior-anchored rating system yielded a current equipment baseline set of scores in the neighborhood of 4 to 5 (adequate to good) on a 7-point rating scale. As employed in this initial research into examining their use as an assessment tool, they did not quite do the intended job. It is concluded that

1. The variation in the scoring was much wider than that required to have the system of scoring be sensitive to the independent variables.

2. At the same time, the distribution of the scoring across the 7-point Likert scale was much narrower than might have been expected, probably because of an inherent tendency of the evaluators to avoid the extreme ends of the rating scale.

3. The purpose of rating with behavioral anchors is to enhance the objectivity of the rating system. Subjectivity remains a significant part of the scoring done in this experiment, in the authors' opinion.

4. The other measures gathered, while forming a somewhat hazy baseline for future assessments, in the case of the CMS-generated data, did not by themselves discriminate among the various independent variables (nor were they necessarily intended to).

5. It is <u>not</u> concluded that the behavior-anchored rating concept for the assessment of ALSE has no merit. It can, in our opinion, be useful if the framework is modified to remove or minimize the conditions that were responsible for generating the imprecision seen in this first research attempt. The following recommendations are made.

RECOMMENDATIONS

The behavioral anchors generated for this experiment should be retained, but they should be reviewed toward editing, to perhaps make them even stronger at each end of their range. The addition of "critical incident" descriptors to their text has been suggested.

The 7-point rating scale should be reduced to either 5 or 3 points. A 3-point scale would simply be 1—poor, unacceptable, a definite decrement in performance relative to current equipment; 2—acceptable, no significant improvement nor decrement in the case of comparisons to current ALSE; 3—good, better than may have been expected, a significant improvement over current equipment. It is believed that this will reduce the standard deviation or variation observed with the 7-point scale.

The evaluators need to learn to be rather ruthless in their scoring of individual performance as affected by the ensemble—no compensating for a subject for any reason. The ALSE is the object of the trials.

It would be desirable to conduct enough future trials to carry encumbered subject aviators to an accommodation asymptote, if there is one. It may be that performance as measured by the behavior-anchored scales, or by any system, would peak and decrease rapidly just because mission flying while encumbered is a very tough job, even in a simulator. Barring such an exercise, continue to adjust evaluation scores for accommodation as was done here.

If a future set of trials is to be conducted to assess prototype or proposed future ensembles or components, the initial series of trials should contain current and proposed ALSE as independent variables in order to perform direct comparative analyses.

Include established techniques for the assessment of situation awareness as it may be influenced by the independent variables.

REFERENCES

Blewett, W.K., Redmond, D.P., Fatkin, L.T., Popp, K., & Rice, D.J. (1995). A P2NBC2 report: Patient decontamination at mobile medical facilities (ERDEC-TR-255). Aberdeen Proving Ground, MD: U.S. Army Chemical and Biological Defense Agency.

Cain Smith, P., & Kendall, L. (1963). Performance appraisal: Assessing human behavior at work In Bernardin & Beatty (1984).

Chatterton, R.T., Vogelsong, K.M., Lu, Y., Ellman, A.B., & Hudgens, G.A. (1996). Salivary amylase as a measure of endogeneous adrenergic activity. Clinical Physiology, 16, 433-448.

Endsley, M.R. (1990). Predictive utility of an objective measure of situation awareness. Proceedings of the Human Factors Society 34th Annual Meeting, p 41-45. Santa Monica, CA: Human Factors Society.

Fatkin, L.T., & Hudgens, G.A. (1994). Stress perceptions of soldiers participating in training at the Chemical Defense Training Facility: The mediating effects of motivation, experience, and confidence level (Technical Report ARL-TR-365). Aberdeen Proving Ground, MD: U.S. Army Research Laboratory.

Headquarters, Department of the Army (1992). Aircrew Training Manual Attack Helicopter, AH-64, (Training Circular TC 1-214). Washington, DC: Author.

Hudgens, G.A., Chatterton, R.T. Jr., Torre, J.P. Jr., Fatkin, L.T., & King, J.M. (1990). Situation characteristics determine stress response profiles. In Proceedings of the Annual Meeting of the American Association for the Advancement of Science. New Orleans:

Kerle, R.H., & Bialek, H. M. (1958). The construction, validation, and application of a Subjective Stress Scale (Staff Memorandum Fighter IV, Study 23). Presidio of Monterey, CA: US Army Leadership Human Research Unit.

Lindquist, E.F. (1953). Design and analysis of experiments in psychology and education. Boston, MA: Houghton Mifflin.

Metzler, T.P. (1995). Presentation to air warrior management working group. St. Louis, MO: Project Manager, Army Command Information System, U.S. Army Aviation and Troop Command, Research, Development, & Engineering Command.

Mullins, L.L., Fatkin, L.T., Patton, D.J. (1998). Soldier perceptions and cognitive performance during physical exertion and heat stress. Unpublished manuscript.

National Research Council (1997). Tactical display for soldiers human factors considerations. Panel on Human Factors in the Design of Tactical Display Systems for the Individual Soldier. Washington DC: National Academy Press.

Reardon, M., Smythe III, N., Hager, J.D., Helms, B., Omer, J., Freeze, M., & Buccannan, D. (1996). Evaluation of the effects of thermally stressful UH-60 simulator cockpit conditions on aviators wearing current MOPP0 and encumbered MOPP IV flight uniforms (ARL TN-96-38). Ft. Rucker, AL: U.S. Army Aeromedical Research Laboratory.

Reardon, M., Smythe III, N., Hager, J.D., Helms, B., Omer, J., Freeze, M., & Buccannan, D. (1996). Physiological and psychological effects of thermally stressful UH-60 simulator cockpit conditions on aviators wearing standard and encumbered flight uniforms (Report No. 97-06). Ft. Rucker, AL: U.S. Army Aeromedical Research Laboratory.

Sharkey, T.J., & Schwirzke, M.F.J. (1995). Investigation of the effects of MOPP IV gear on flight performance in the crew station research and development facility (CSRDF) flight simulator (Report Number MTI-TR-95-940506(11)-01). Moffet Field, CA: Monterey Technologies, Inc., for USAATCOM Aeroflight Dynamics Directorate.

Shively, J., Atencio, A., Shankar, R., Bunzo, M., Bushnell, H., Neukom, C., Wong, C., & Banda, C. (1995). MIDAS evaluation of AH-64D Longbow Apache crew procedures in an air-ground flight segment: MOPP versus unencumbered Moffet Field, CA: USAATCOM RDEC Aeroflight Dynamics Directorate.

Simon, R. (1991). Results of the data analysis army aircrew coordination measures testbed conducted in spring 1990. Wilmington, MA: Dynamics Research Corporation.

Simon, R., Grubb, G., & Leedom, D. (1992). Development of crew coordination evaluation methods and materials (DRC Report E-21455U). Wilmington, MA: Dynamics Research Corporation.

Simon, R., Grubb, G., & Leedom, D. (1993). Development of candidate crew coordination evaluation methods and materials (DRC Report E-21867U). Wilmington, MA: Dynamics Research Corporation.

Tauson, R.A., Doss, N.W., Rice, D.J., Tyrol, D.E., & Davidson, D. (1995). The effect of vehicle noise and vibration (caused by moving operations) on cognitive performance in the command and control vehicle (Technical Report ARL-MR-279). Aberdeen Proving Ground, MD: U.S. Army Research Laboratory.

Torre, J.P., Wansack, S., Hudgens, G.A., King, J.M., Fatkin, L.T., Mazurczak, J., & Breitenbach, J.S. (1991). Effects of competition and mode of fire on physiological responses, psychological stress reactions, and shooting performance (Technical Memorandum 11-91). Aberdeen Proving Ground, MD: U.S. Army Research Laboratory.

U.S. Army Aviation Center (1992). Crew coordination exportable evaluation package for Army aviation. Ft. Rucker AL: Prepared by Army Research Institute for the Behavioral Sciences and Dynamics Research Corporation.

U.S. Army Materiel Command (1996). Instructor utilization handbook AH-64 combat mission simulator (Revision A). Orlando FL: Simulation, Training, and Instrumentation Command.

Urwick, L., & Brech, E.F.L. (1965). The making of scientific management: Vol 3. The Hawthorne investigations. London: Pitmans.

Wright, R.H., Hartson, R.R., & Couch, M.E. (1996). Air warrior baseline evaluation, Vol. I, Summary. Ft. Rucker, AL: U.S.Army Research Institute for the Behavioral Sciences.

Ylonen, H., Lyytinen, H., Leino, T., Leppaluoto, J., & Kuronen, P. (1997). Heart rate responses to real and simulated BA Hawk MK 51 flight. Aviation Space and Environmental Medicine, 68, 601-605.

Zuckerman, M., & Lubin, B. (1985). Manual for the multiple affect adjective checklist—revised. San Diego, CA: Educational and Industrial Testing Service.

APPENDIX A

DESCRIPTION OF ALSE COMPONENTS

DESCRIPTION OF ALSE COMPONENTS

Aircraft: AH-64

Crew Position: Pilot-Front seat, Pilot-Gunner Rear seat

Mission profile: Temperate environment, Combat, Over water, NBC Threat

1. Coverall: Battle Dress Overgarment, NBC or Coverall CWU-27P or CWU-21A/P
2. Boots: Leather jump boot; Danner; MIL-B-44152
3. Overshoes: NBC Rain boot
4. Gloves: Std. nomex aircrew MIL-G-81188
5. Overgloves: Butyl Rubber NBC
6. Mask: M43 Type I
7. Helmet: Integrated Helmet And Display Sight System (IHADSS)
8. Side arm: M-9 Beretta, holster
9. Survival Vest
10. Floatation vest: Preserver, life, underarm, LPU-10/P
11. Underwater breathing: HEED SRU-36/P
12. Raft: SRU-37P (worn anterior)
13. Body Armor: Front (only) chicken plate
14. Radio: Survival AN/PRC-112A

Order: Coverall, Survival vest, Floatation vest, Raft, Body armor

APPENDIX B

SURVEYED ATM TASKS

SURVEYED ATM TASKS

TASK No.	Category	Task Name
1106	Front Seat	Perform data entry procedures
2044	Emerg Pro	Perform actions on contact
1101	Front Seat	Perform TADS boresighting
1051	Emerg Pro	Perform standard autorotation
1145	Both/either	Perform IHADSS target tracking
2049	Front Seat	Search for and identify targets with the TADS
1037	Terrain Flt	Perform NOE deceleration
1100	Front Seat	Perform TADS operational checks
1035	Terrain Flt	Perform terrain flight
1054	Emerg Pro	Perform simulated single-engine failure, OGE hover
1017	Hover	Perform hovering flight
1103	Both/either	Perform IHADSS boresighting
1140	Both/either	Engage target with Hellfire
1141	Both/either	Engage target with the ARCS
2052	Front Seat	Perform target tracking with the TADS
1026	Terrain Flt	Perform doppler navigation
1062	Emerg Pro	Perform ECU lockout operations
1142	Both/either	Engage target with the AWS
1081	TO/Ldg	Perform nonprecision approach
1032	Hover	Perform slope operations
1052	Emerg Pro	Perform simulated engine failure, IGE hover
1055	Emerg Pro	Perform single-engine landing
1079	Both/either	Perform radio communication procedures
1090	Hover	Perform masking and unmasking
1098	Back Seat	Perform after-landing tasks
1031	TO/Ldg	Perform confined area operations
1078	Emerg Pro	Perform unusual attitude recovery
1007	Both/either	Perform engine-start, run -up, hover, and before-takeoff checks
1019	TO/Ldg	Perform a rolling takeoff (minimum power takeoff)
1020	TO/Ldg	Perform simulated maximum
1015	TO/Ldg	Perform ground taxi
1039	Cruise Flt	Perform high-speed flight
1028	TO/Ldg	Perform VMC approach
1018	TO/Ldg	Perform a normal takeoff
1021	Cruise Flt	Perform deceleration/acceleration

APPENDIX C

HASTY ATTACK INTO HORSE

AIRCREW MISSION BRIEFING

For use of this form. See AR 95-1 and Unit SOP.____________

UNIT: _A Co._______ DATE: PRESENT MISSION NO.: ___HAHOR-IC-128/TEE-516_

BRIEFING: See FRAGO__ See FRAGO

1. ENEMY: _HAHOR01___ FRIENDLY: _HAHOR01_ TASK ORG: X CORPS

*2. MISSION: _Conduct a hasty attack to destroy enemy tank and _______ _motorized rifle units in the vic of EA HORSE.___________________

3. EXECUTION:
   - *a. Mission Type: TAC TNG ADMIN MAINT MED OTHER
   - *b. Authorized Conditions: DAY N VMC IMC NVD HOOD OTHER
   - *c. Authorized Flight Modes: Form Low-level Contour NOE
   - *d. Aircraft/crews/duty status:

	A/C TYPE/TAIL #	PC/SEAT	PI/SEAT
(1)	_AH-64 /G16______	POTTS/ BS______	HODGKINS/ FS_____
(2)	_AH-64 /G12______	__CMS CREW_____	
(3)			
(4)			
(5)			

   - *e. Special mission equipment: Operational ALQ-136.144 and APR 39
   - f. Authorized Loads: PAX N/A Cargo/Sling N/A AMMO SEE REMKS
   - *g. Flight Route: O?O. Depart HA Astros along Air Axis OWL to____ occupy BPs O01 and O-2. Maintain 50 ktas and 100 ft AGL______ until 2 kms from BP. Formation - combat cruise. Return same_ route 120 ktas and 100 ft AGL.________________________________
   - *h. Mission Restrictions: _Do not go east of N/S Grid 78_________
   - *i. Safety Considerations: ID all targets. 8/17 CAV in VIC of BPS_
   - *j. Risk assessment for this mission is: HIGH MEDIUM LOW

4. SERVICE SUPPORT:
   - a. Refuel/Rearm location: PRI-MURPH VK 7052 6940/SEC-JUMP VK 7460 6645
   - b. Assembly Area/Bivouac/RON Locations: FAA BRAVES VK 8785 8091____
   - c. Maintenance Support: Limited-HA JUDY and ASTROS - Full Cap.-HA BETH

5. COMMAND AND SIGNAL:
   - *a. Command:
     - (1) Air Mission Commander: _CPT POTTS________________________
     - (2) Command or Support Relationship: 8/17 Attached X CORPS OPCON
   - b. Signal (except published Freq.) See Commo Card. Report BP arrival and departure. Report passing all Phase Lines. and report FARP___ arrival and departure to 8/17th Squadron S-3.____________________

6. ADDITIONAL REMARKS: LASER CODES: G-12: A (1112). G16: b (1116)___ 8/17 CAV GLLD: C28: C (1128)____

   Weapons Load: 1200 RDS 30 MM. 8 HELLFIRE. 38 ROCKETS = E ZONE: 6MP

__/s/ Homer S. Reno_

BRIEFER’S SIGNATURE

MAJ, AVN

S-3

*** MANDATORY FOR ALL FLIGHTS**

Copy_____ of _____ Copies
TF DIGGER,Western Coalition
oyal Moldovian Kingdom
051400SJUN9__
MAJ Payne

FRAGO HAHOR01

Reference: Simulator Map 2317 series M741S “TROUT LAKE” 1:50,000
Time Zone used throughout Sierra (Local)

1. SITUATION:

a. Enemy: Early this morning Motorized Rifle and Tank Forces of the Peoples Krasnovian Republic launched a major attack along the southern border of the Moldovian Kingdom. The main effort of the attack is at Avenue of Approach B and C. Currently, lead elements of the 9th GMRR and 21st MRR are being heavily engaged by units of the 38th and 80th ID in vic of EA BIRD and DOG. A faint attack by the 3 MRR into the 2nd BDE of the 36th ID operational area has caused a shift of forces to the west creating a large gap between the 1st and 2nd BDE of the 80th ID. A Tank BN and Motorized Rifle BN Task force has broken away from the 21st MRR and attempting to split the 1st and 2nd BDE and gain access to the Division Rear.

b. Friendly: 1st BDE 80th ID has shifted west, closer to the western boundary of the Division and the 2nd BDE has shifted to protect the Division Center and Eastern boundary. This repositioning has caused a major gap between the 1st and 2nd BDE and has thin the 2nd BDE lines. The 38th ID has counter attacked and pushed the 21st MRR just south of EA BIRD and now are in defensive positions along the East/West grid 59. Units of the 80th ID are still in a pitched battle along the East/West grid 60 with heavy casualties to both sides. The 3rd BDE 80th ID is being held in reserve by Corps and mechanized infantry units of the 8/17 Cav are being rushed to the vic EA HORSE.

2. MISSION: OPCON to the 8/17th Cav will conduct a hasty attack and destroy enemy forces in EA HORSE.

3. EXECUTION: A co 1-633 ATKHB moves to contact along Air Axis OWL as soon as possible and attack motorized rifle and tank units in EA HORSE from BPs 0-1 and 0-2.

4. Service Support: No change

5. Command and Signal: No change

6. WEATHER: Clear and 15 NM Baro Press: 30:00
TEMP: 15 C PA: 438

BENTEEN
COL
8/17th CAV

Official:
RENO
S3
Annex: A X Corps and 8/17th Cav OPNS Graphics

APPENDIX D

SAMPLE OF BEHAVIORAL ANCHORS AND GRADE SLIP

SAMPLE OF BEHAVIORAL ANCHORS AND GRADE SLIP

Basic Qualities and Behavioral Anchors
for the Assessment of Aircrew Life Support Equipment (ALSE)
as it Affects Aircrew Performance

Overview

This method of assessment of ALSE is based on the highly successful Crew Coordination Training program instituted by the U.S. Army Aviation Center (USAAVNC) and the *Crew Coordination Exportable Evaluation Package for Army Aviation* developed by the Army Research Institute for the Behavioral Sciences. The latter will serve as a framework for the evaluation of aircrew behaviors as influenced by, and aircrew performance as may be degraded by the wearing of ALSE during operations. The purpose of conducting such assessments and evaluations is to: (1) establish baseline bodies of data reflecting the expected performance, measures of effectiveness and qualities of mission operations employing the current collection of ALSE components and (2) conduct comparative assessments of developing ALSE which may take advantage of emerging technologies and the initiative to more thoroughly integrate ALSE components to each other, the aircraft and most importantly, the Air Warrior.

The <u>Basic Quality</u> items below are the behavioral anchors to be employed in the assessment. They have been adapted for the evaluation of behaviors affect by ALSE by modification and editing of the Crew Coordination Evaluation items as they appear in the Exportable Package. It was the intent of the research investigators to retain the crew coordination qualities of the items as the underlying basis of the assessment methodology, but to focus on the affect of ALSE as reflected in aircrew behaviors. In-other-words, the aviators both individually and collectively as the aircrew, will have become the indicators, the yardstick, employed to measure the influence and ultimate acceptability of ALSE in the cockpit.

Personnel trained and experienced in conducting Crew Coordination Evaluations should, with a minimum of additional training, be able to adapt and serve as evaluators for this Air Warrior project. In performing mission flight, the primary considerations that the evaluators must keep in mind are summarized in the following questions:

- Is the aircrew able to overcome the influence or affect of the ALSE in terms of crew coordination and performance of tasks, or is their behavior, in fact, degraded because of the ALSE?

- Is the aircrew mentally preoccupied or distracted by the ALSE during the mission, which, in turn, is hurting their crew coordination and performance?

- Are members of the aircrew physically hampered in crew coordination efforts and the performance of tasks because of interferences between the ALSE, the aircraft and their own bodies?

==
BASIC QUALITY 1. Establish and maintain flight team leadership and crew climate (Crew Climate)
==
Explanation:

This rating assesses the quality of relationships among the crew and the overall climate of the flight deck.
Aircrews are teams with a designated leader and clear lines of authority and responsibility. The pilot-in-command sets the tone of the crew and maintains the working environment. Effective leaders use their authority but do not operate without the participation of other crewmembers. When crewmembers disagree on a course of action, rate the crew's effectiveness in resolving the disagreement. Note: Traditional leadership centralizes leadership in the leader with followers fully dependent on the leader. Functional leaderships assigns leadership and followership roles as the situation evolves. Flight team leadership recognizes the impact of leadership style on the working environment. Regardless of leadership style, the pilot-in-command retains final decision and direction authority.

Superior Rating (7)

The crewmembers have very good interpersonal relationships, with no detriment from ALSE caused stressors. They respect each others' skills and appear to enjoy being with each other. The climate is very open; crewmembers freely talk and ask questions. Crewmembers encourage the individual with the most information about the situation-at-hand to participate. There is a genuine concern for good working relationships. No degrading comments or negative voice tones are used in interactions. Disagreements are perceived as a normal part of crew interactions, and the crew directly confronts the issues over which the disagreement began. Arguments or disagreements focus on behaviors or solutions, rather than on personalities. Disagreements do not arise due to irritation and frustration created by ALSE caused stressors. Each crewmember carefully listens to others' comments. Senior crewmembers accept challenges from junior crewmembers. Alternative solutions are explored. The solution produced is a "win-win" situation in which all crewmembers' opinions are considered. The crewmembers have no hard feelings at the conclusion of the incident.

Acceptable Rating (4)

The crewmembers have sound interpersonal relationships and seem to respect each others' skills. The climate is an open one, in which ALSE stressors play little or no role, and crewmembers are free to talk and ask mission questions. Regardless of rank or duty position, the individual with the most information about the situation-at-hand is allowed to participate. When disagreements arise, it may be because of the ALSE. The crew directly confronts the issues over which the disagreements began. The primary focus is on behaviors or solutions not on ALSE, and no personal attacks are made in the heat of discussion. The solution is generally seen as reasonable. Problem resolution ends on a positive note with very little hostility or grumbling among crewmembers. Mutual respect is clearly intact.

Very Poor Rating (1)

Crew interactions are often awkward and uncomfortable due to ALSE caused stressors. The crewmembers do not appear to like or respect each other. Crewmembers may be curt and impolite to each other as a result of ALSE stressors. Requirements for assistance are made as commands rather than as requests for support. When problems arise, the crew fails to directly confront the issues due to distraction or preoccupation with the ALSE and fail to recognize that the ALSE may be the underlying cause. Personal attacks may arise. Senior crewmembers are resistant to recommendations from junior crewmembers. Crewmembers do not explore the range of possible solutions. They may shout and argue without finding a solution. One or more crewmembers may retreat and say nothing at all. A “win-lose” situation develops in which one crewmember is shown to be right and the other to be wrong. The crewmembers show little respect to one another except for deferring to formal rank. All such behaviors are attributed to the ALSE.

==

BASIC QUALITY 8. Decisions and actions communicated and acknowledged (Comm/Act)

==

Explanation:

Rate the extent to which decisions and actions are actually made and announced to the crewmembers after input is solicited from them. Crewmembers should respond verbally or with the appropriate adjustment to their behaviors, actions, or control inputs to clearly indicate that they understand when a decision has been made and what it is. Failure to do so may confuse crews and lead to uncoordinated operation. Note: Due to time constraints in certain situations, there is often little or no more time for crews to make inputs to a decision. In such cases, raters should focus on the extent to which decisions are acknowledged verbally or through coordinated, pre-planned action.

Examples:

- UH-60 Task 2086, Perform masking and unmasking: P* will announce his intent to unmask. The P and CE will acknowledge that they are prepared to execute the maneuver.

- AH-64 Task 1038, Perform terrain flight approach: P* will announce intention of a go-around...whether approach will terminate to a hover or to the ground. P will acknowledge use of manual stabilator or any intent to deviate from the approach.

Superior Rating (7)

The pilot-in-command states, without interference of or preoccupation with ALSE, decisions and actions and, time permitting, explains the reasons and intent. Crewmembers acknowledge the decisions with a clear verbal response, without ALSE interference and ask questions to clarify any confusion. The pilot-in-command answers all questions in a positive, straight-forward, easy to hear manner. Crewmembers keep the pilot-in-command informed of the results of their activities and changing responsibilities - especially visual area of responsibility or task focus, as may be influenced by ALSE. The crew clearly acknowledges, without ALSE influence, results of actions, or changes, and then states its intended adjustments based on the information provided. If crewmembers do not acknowledge or adjust, the pilot-in-command requests acknowledgment. Crewmembers are particularly attentive to the communication of workload, stress, heat stress etc. buildups. When assuming control of the aircraft or making control inputs, notification is always given and acknowledgment received regardless of ALSE interference.

Acceptable Rating (4)

The pilot-in-command states, with a minimum of interference of or preoccupation with ALSE, decisions and actions along with, time permitting, a brief explanation of the reasons and informs the crew of the adjustments they are expected to make. The crew acknowledges, without ALSE influence, its awareness of the decisions and directions. Crewmembers may ask questions to clarify confusion. The pilot answers questions clearly, with little ALSE or speech and hearing interference, and quickly and the crew adjusts to the new situation. When assuming control of the aircraft or

making control inputs, notification is given and acknowledged with a minimum of ALSE interference.

Very Poor Rating (1)

Decisions and actions of a crewmember are often not passed on to the crew, due to interference of or preoccupation with ALSE. The pilot-in-command takes unilateral action and can not clearly explain or inform the crew of his intended purpose due to ALSE influences, like poor speech intelligibility. The crew is often not aware that a decision has been made. The crew infrequently asks questions for clarification. The pilot-in-command may not acknowledge or respond to questions. The crew may not know how to react to changed circumstances. Crewmembers may take uncoordinated actions without being able to clearly state intentions or results, due to ALSE induced confusion. Two pilots may attempt to simultaneously take control of the aircraft, because of degraded speech intelligibility and poorer hearing ability, when flight control authority is unclear. Neither crewmember may be clearing the aircraft.

==
BASIC QUALITY 12. Advocacy and assertion practiced (Advoc/Assert)
==

Explanation:

This rating evaluates the extent to which crewmembers advocate a course of action they consider best, when it may differ with the one being followed or proposed. Note: Except under extreme emergency conditions where time is absolutely critical, it is usually in the crew's best interest to hear the full range of viewpoints available.

Examples:

- UH-60 and AH-64 Task 2083, Negotiate wire obstacles: Crew will discuss the characteristics of the wires...to determine the method of crossing.

- AH-64 Task 2044, Perform actions on contact: Crew will discuss options for developing the situation.

Superior Rating (7)

Crewmembers clearly state to the rest of the crew a course of ALSE influenced action that they consider best. They clearly explain their reasons for believing this to be the best course. Other crewmembers listen to the argument before presenting any criticism or proposing alternate ALSE based actions. Discussions focus on the strengths and weaknesses of the proposed course of ALSE based action, without lengthy discourse on ALSE caused stressors. Crewmembers call the crew's attention to changes in the situation, as may be influenced by ALSE, and provide information that is essential to the proper execution of another crewmember's task. Crewmembers pursue feedback to ensure that their views are heard and understood. Other crewmembers expect such open comments and view them as positive contributions to mission performance in spite of ALSE.

Acceptable Rating (4)

Crewmembers state their support for a course of action or suggest improvements to other proposed ALSE influenced actions. Each crewmember makes an effort to explain his position and convince others to concur with him on the course of the ALSE based action to be taken. Other crewmembers may interrupt with their views and alternatives. Crewmembers usually speak out when they recognize a departure from the best plan, as influenced by ALSE, or standard procedures or when they have a piece of information that is important to another's task execution, e.g., using the HDV and not the ORT. Crewmembers seek assurances that presented information has been received. Other crewmembers view such comments as constructive and not a challenge to authority.

Very Poor Rating (1)

The crew almost never suggest course of ALSE influenced action. Crewmembers attempting to propose a course of action may be cut-off before they can propose the ALSE based action or explain the rationale for that action. Crewmembers proposing courses of action, as influenced by ALSE, may receive personal attacks due to frustration caused by ALSE stressors like heat build-up or confinement. The crew raises few, if any ALSE concerns. Crewmembers may even fail to intervene when ALSE caused risks arise.

APPENDIX E

AIRCREW COORDINATION EVALUATION GRADE SLIP

AIRCREW COORDINATION EVALUATION GRADE SLIP

AIRCREW COORDINATION EVALUATION (ACE) CHECKLIST

For use of this form, see Aircrew Coordination Exportable Evaluation Package for Army Aviation.

PC ____________________ Date __________

PI ____________________

NCM ____________________

____

NO	CREW COORDINATION BASIC QUALITIES	RATING
1	Establish and maintain flight team leadership and crew climate (Crew Climate)	
2	Premission planning and rehearsal accomplished (Plan & Rehearse)	
3	Application of appropriate decision making techniques (Decision Tech)	
4	Prioritize actions and distribute workload (Workload)	
5	Management of unexpected events (Unexp Events)	
6	Statements and directives clear, timely, relevant, complete, and verified (Info Xfer)	
7	Maintenance of mission situational awareness (Sit Aware)	
8	Decisions and actions communicated and acknowledged (Comm/Ack)	
9	Supporting information and actions sought from crew (Info Sought)	
10	Crewmember actions mutually cross-monitored (Cross-Monitor)	
11	Supporting information and actions offered by crew (Info Offered)	
12	Advocacy and assertion practiced (Advoc/Assert)	
13	Crew-level after-action reviews accomplished (AAR)	

Evaluator's Signature.

Notes:
Consult the behavioral anchored rating guidance. Enter a summary rating (1, 2 .. 7) in the rating block for each Basic Quality. Refer to the rating scale below.

RATING SCALE

Very Poor	Poor	Marginal	Acceptable	Good	Very Good	Superior
1	2	3	4	5	6	7

AIRCREW COORDINATION EVALUATION (ACE) CHECKLIST

NO. OF COPIES	ORGANIZATION
2	ADMINISTRATOR DEFENSE TECHNICAL INFO CENTER ATTN DTIC OCP 8725 JOHN J KINGMAN RD STE 0944 FT BELVOIR VA 22060-6218
1	DIRECTOR US ARMY RESEARCH LABORATORY ATTN AMSRL CS AL TA REC MGMT 2800 POWDER MILL RD ADELPHI MD 20783-1197
1	DIRECTOR US ARMY RESEARCH LABORATORY ATTN AMSRL CI LL TECH LIB 2800 POWDER MILL RD ADELPHI MD 207830-1197
1	DIRECTOR US ARMY RESEARCH LABORATORY ATTN AMSRL D DR R WHALIN 2800 POWDER MILL RD ADELPHI MD 20783-1197
1	DIRECTOR US ARMY RESEARCH LABORATORY ATTN AMSRL DD J J ROCCHIO 2800 POWDER MILL RD ADELPHI MD 20783-1197
1	WALTER REED ARMY INST OF RSCH ATTN SGRD UWI C (COL REDMOND) WASHINGTON DC 20307-5100
1	COMMANDER USA AEROMEDICAL RESEARCH LAB ATTN LIBRARY FORT RUCKER AL 36362-5292
1	CHIEF ARMY RESEARCH INSTITUTE AVIATION R&D ACTIVITY ATTN PERI IR FORT RUCKER AL 36362-5354
1	MR R BEGGS BOEING-HELICOPTER CO P30-18 PO BOX 16858 PHILADELPHIA PA 19142

NO. OF COPIES	ORGANIZATION
1	DIRECTOR US ARMY AEROFLIGHT DYNAMICS DIR ATTN SAVRT AF D (A W KERR) AMES RESEARCH CENTER (MS 215-1) MOFFETT FIELD CA 94035-1099
1	PROGRAM MANAGER RAH-66 ATTN SFAE AV BLDG 5300 SPARKMAN CENTER REDSTONE ARSENAL AL 35898
1	DENNIS L SCHMICKLY CREW SYSTEMS ENGINEERING MCDONNELL DOUGLAS HELICOPTER 5000 EAST MCDOWELL ROAD MESA AZ 85205-9797
1	JON TATRO HUMAN FACTORS SYSTEM DESIGN BELL HELICOPTER TEXTRON INC PO BOX 482 MAIL STOP 6 FT WORTH TX 76101
1	CHIEF CREW SYSTEMS INTEGRATION SIKORSKY AIRCRAFT M/S S3258 NORTH MAIN STREET STRATFORD CT 06602
1	COMMANDER US ARMY AVIATION CENTER ATTN ATZQ CDM S (MR MCCRACKEN) FT RUCKER AL 36362-5163
1	DIRECTOR US ARMY AEROFLIGHT DYNAMICS DIR MAIL STOP 239-9 NASA AMES RESEARCH CENTER MOFFETT FIELD CA 94035-1000
2	DIRECTOR ARMY NATIONAL GUARD ATTN NGB AVN MAJ WEST LTC BACON 111 SOUTH GEORGE MASON DRIVE ARLINGTON VA 22204
1	CDR U S ARMY AVIATION TECHNICAL TEST CENTER ATTN STEAT CO CAIRNS ARMY AIRFIELD BLDG 30601 FT RUCKER AL 36362

NO. OF COPIES	ORGANIZATION
1	DIRECTOR DIRECTORATE OF COMBAT DEVELOPMENTS US ARMY AVIATION CENTER ATTN ATZQ CD (MR G HARRISON) FORT RUCKER AL 36362
1	DIRECTOR DIRECTORATE OF EVALUATION AND STANDARDIZATION US ARMY AVIATION CENTER ATTN ATZQ DPT ES FORT RUCKER AL 36362
1	US ARMY RESEARCH INSTITUTE ROTARY WING AVIATION RSCH UNIT ATTN TAPC ARI IR (DR R WRIGHT) FORT RUCKER AL 36362-5000
1	COMMANDING GENERAL US ARMY AVIATION CENTER ATTN ATZQ TSM LB FORT RUCKER AL 36362-5000
1	DIRECTOR DIRECTORATE OF TRAINING DOCTRINE & SIMULATION ATTN ATZQ TDS GT (MS L THOMPSON) FORT RUCKER AL 36362
1	HHC (USAG) ATTN AFZF AV ST (CW5 T CONNELL) FT HOOD TX 76544
1	PROGRAM MANAGER AIR CREW INTEGRATED SYSTEMS ATTN SFAE AV LSE (MR T METZLER) BLDG 5681 REDSTONE ARSENAL AL 35898
1	CDR ARMY AVIATION SUPT FACILITY #2 ATTN MAJ K NETTLES SILVER BELL ARMY HELIPORT AZ 85653
1	CDR WESTERN ARNG AVTN TRAINING SITE SILVER BELL ARMY HELIPORT AZ 85653-9598
1	CDR WESTERN ARNG AVTN TRAINING SITE AVTN SUPPORT & TRAINING TASK FORCE ATTN LTC J BRAMAN SILVER BELL ARMY HELIPORT AZ 85653-9598
1	CDR WESTERN ARNG AVTN TRAINING SITE AVTN SUPPORT & TRAINING TASK FORCE ATTN MAJ D MITCHELL SILVER BELL ARMY HELIPORT AZ 85653-9598
	ABERDEEN PROVING GROUND
2	DIRECTOR US ARMY RESEARCH LABORATORY ATTN AMSRL CI LP (TECH LIB) BLDG 305 APG AA
1	LIBRARY ARL BLDG 459 APG-AA
1	USATECOM RYAN BUILDING APG-AA
1	CHIEF ARL HRED ERDEC FIELD ELEMENT ATTN AMSRL HR MM (R MCMAHON) BLDG 459 APG-AA
1	US ARMY EDGEWOOD RDEC ATTN SCBRD TD J VERVIER APG MD 21010-5423
	ABSTRACT ONLY
1	DIRECTOR US ARMY RESEARCH LABORATORY ATTN AMSRL CS AL TP TECH PUB BR 2800 POWDER MILL RD ADELPHI MD 20783-1197

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
March 1999

3. REPORT TYPE AND DATES COVERED
Final

4. TITLE AND SUBTITLE
Aviator Behavior and Performance as Affected by Aircrew Life Support and Protective Equipment

5. FUNDING NUMBERS
AMS Code 611102.74A00011
PR: 1L1611102.74A
PE: 6.11.10

6. AUTHOR(S)
Waugh, J.D.; Fatkin, L.T.; Patton, D.J.; Mullins, L.L.; Burton, P.A.; Barker, D.J. (all of ARL); Mitchell, D.A. (Arizona ANG)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Research Laboratory
Human Research & Engineering Directorate
Aberdeen Proving Ground, MD 21005-5425

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Laboratory
Human Research & Engineering Directorate
Aberdeen Proving Ground, MD 21005-5425

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
ARL-MR-440

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

A methodology for quantifying Army rotary wing aviator performance as influenced by aircrew life support, survival, and nuclear-biological-chemical clothing and equipment ensembles was examined in a set of experimental trials conducted in an AH-64 (Apache) combat mission simulator. The methodology was based on an aircrew evaluation procedure originally developed for use in the crew coordination training of all Army aviators. It uses a set of 13 basic qualities, each with behavioral anchors and a 7-point rating scale, and it is administered by specifically trained senior aviator evaluators. Ten crews, two aviators in each, while fully encumbered, performed three combat missions for record, representative of typical operational tasks, with one "variation" trial conducted without the over-water components of the ensemble. Measures of effectiveness and flight data, as well as stress assessment and equipment "complaints" citations, were recorded. The results indicated that the behavior-anchored scores were not sensitive enough to statistically discriminate among the independent variables of repeated measures and the variation trials even though graphically, differences were readily apparent. Attempts to apply transformations to the data, based on the aviator subjects' relative flying experience and their apparent accommodation to the trials were also statistically unsuccessful. The additional measures collected did not yield statistically significant discriminations nor did they correlate well with the evaluation scores. A number of options for improving the technique are offered.

14. SUBJECT TERMS

aviator performance	Apache attack helicopter	life support equipment	survival equipment
air crew integrated systems	behavioral anchors	NBC ensemble	
air warrior	crew coordination evaluation	rotary wing aviation	

15. NUMBER OF PAGES
120

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT
Unclassified

18. SECURITY CLASSIFICATION OF THIS PAGE
Unclassified

19. SECURITY CLASSIFICATION OF ABSTRACT
Unclassified

20. LIMITATION OF ABSTRACT